

FIG. 1A

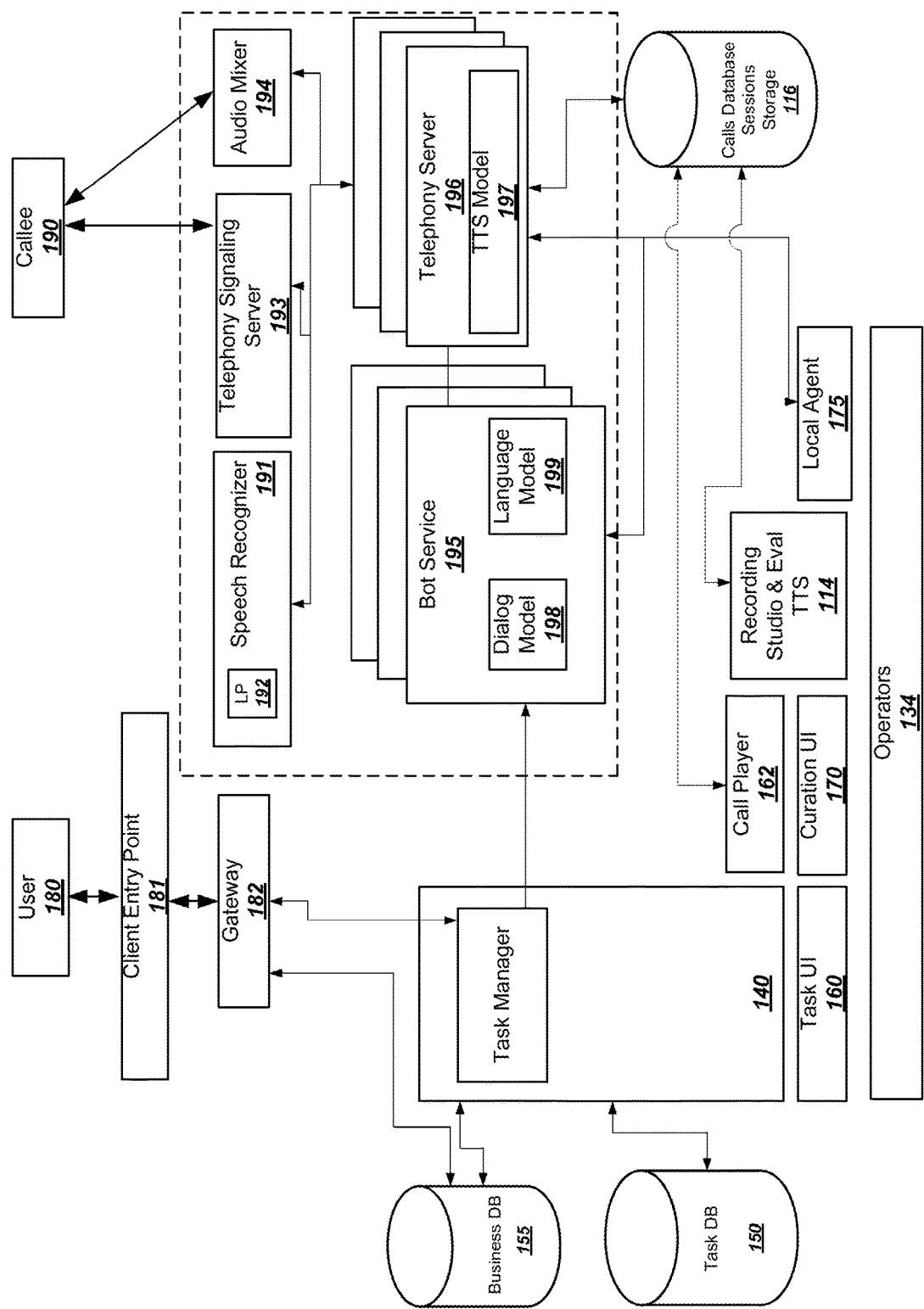


FIG. 1B

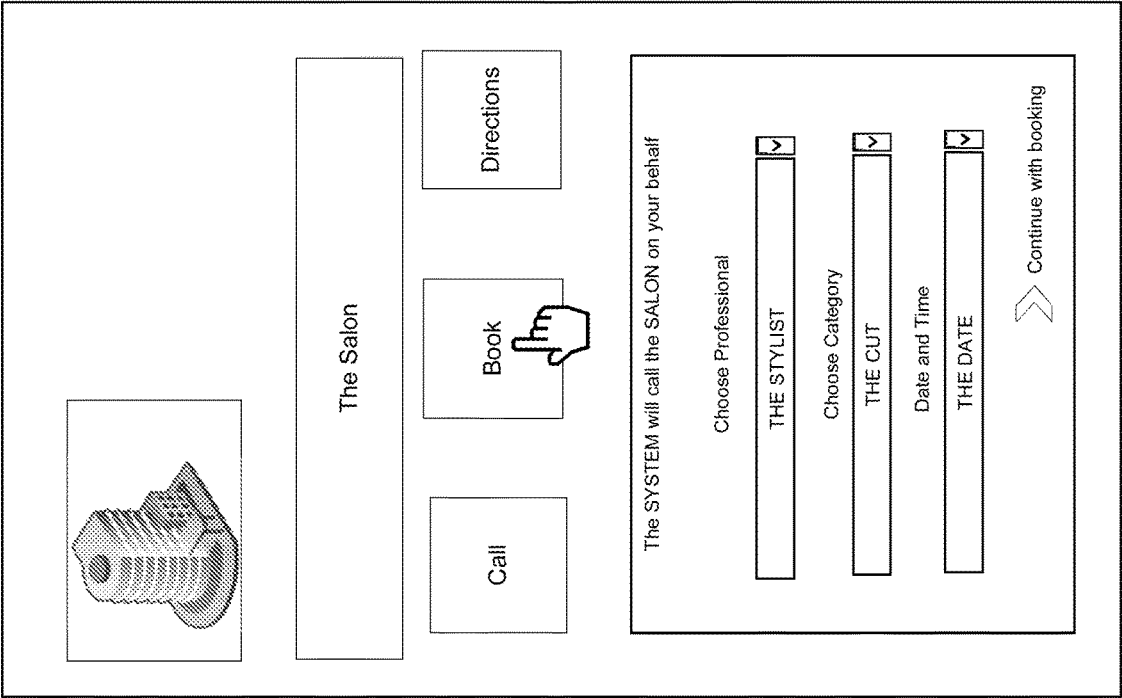


FIG. 1C

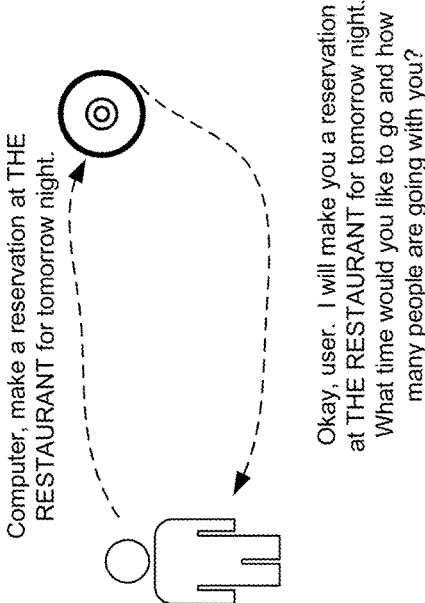


FIG. 1D

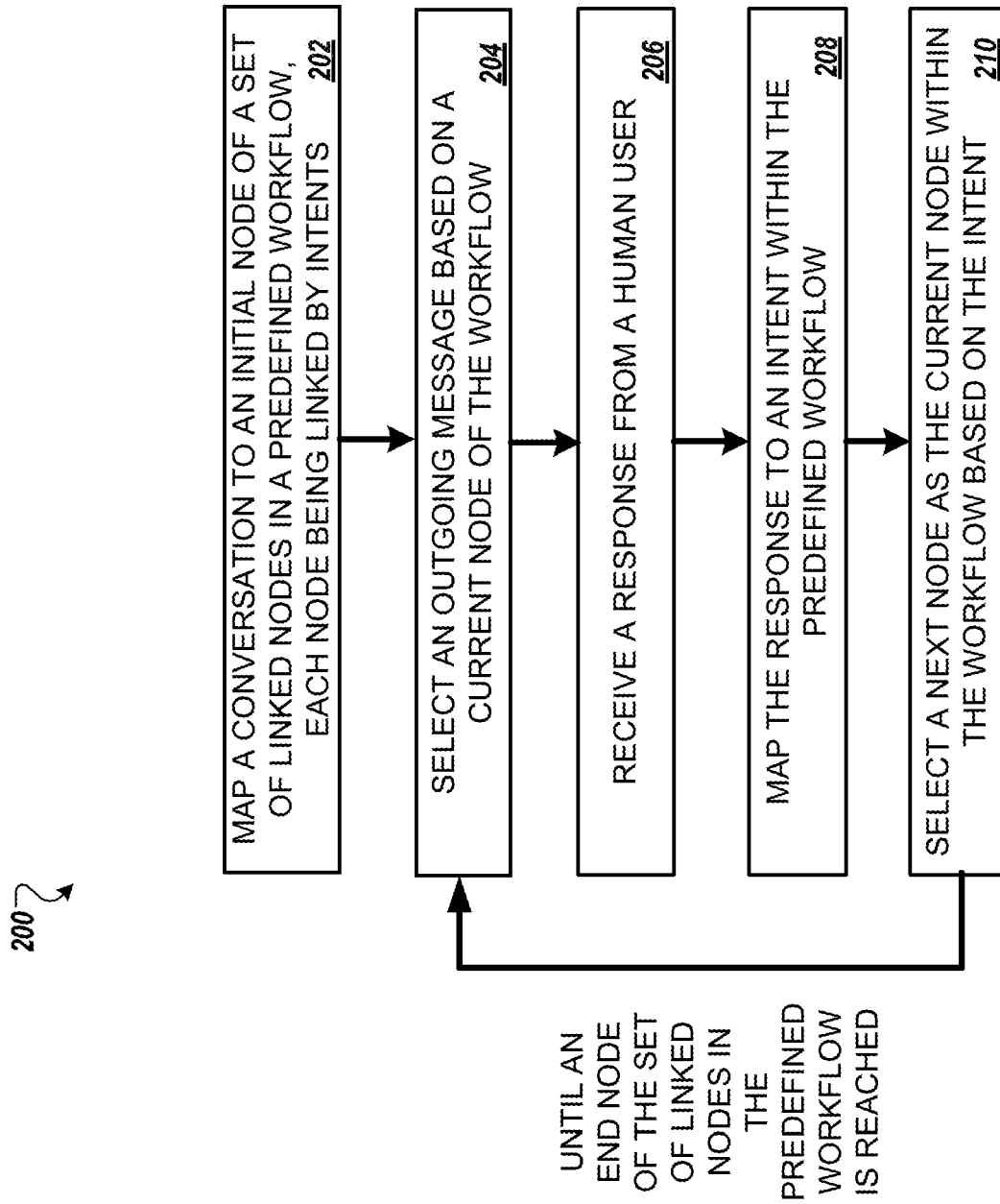


FIG. 2A

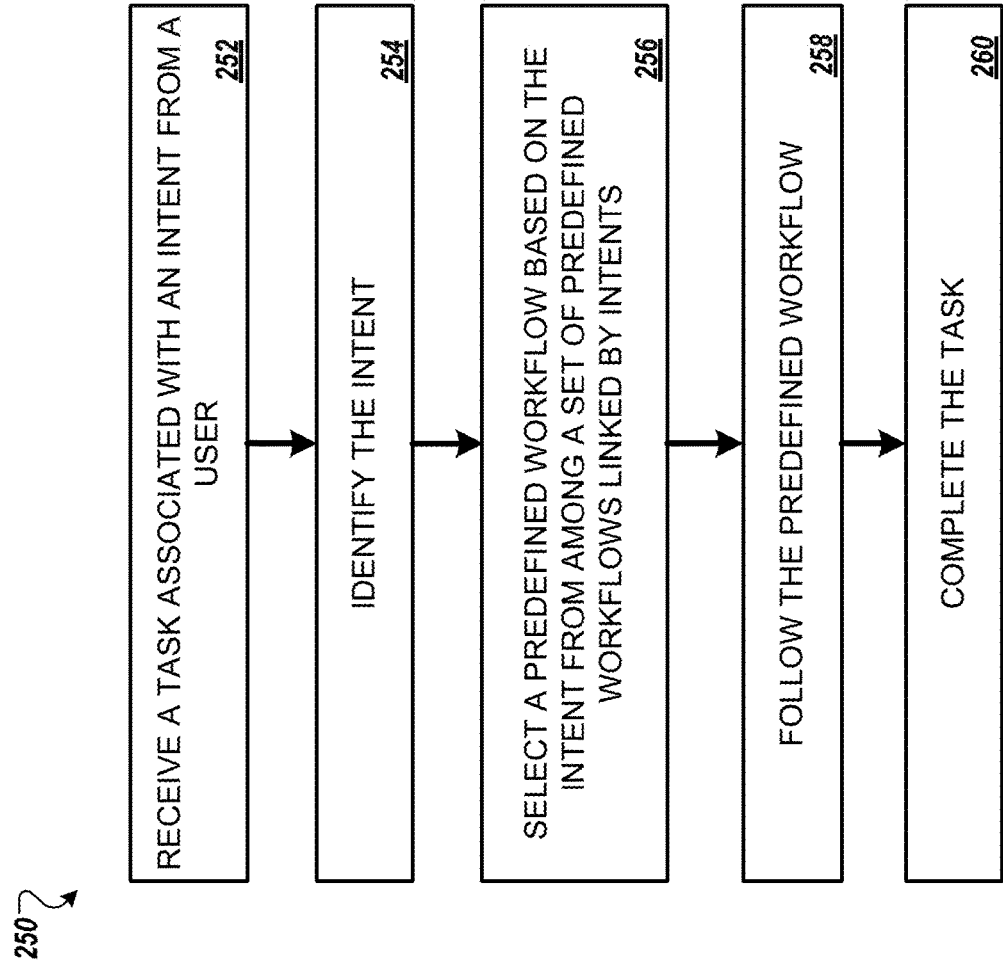


FIG. 2B

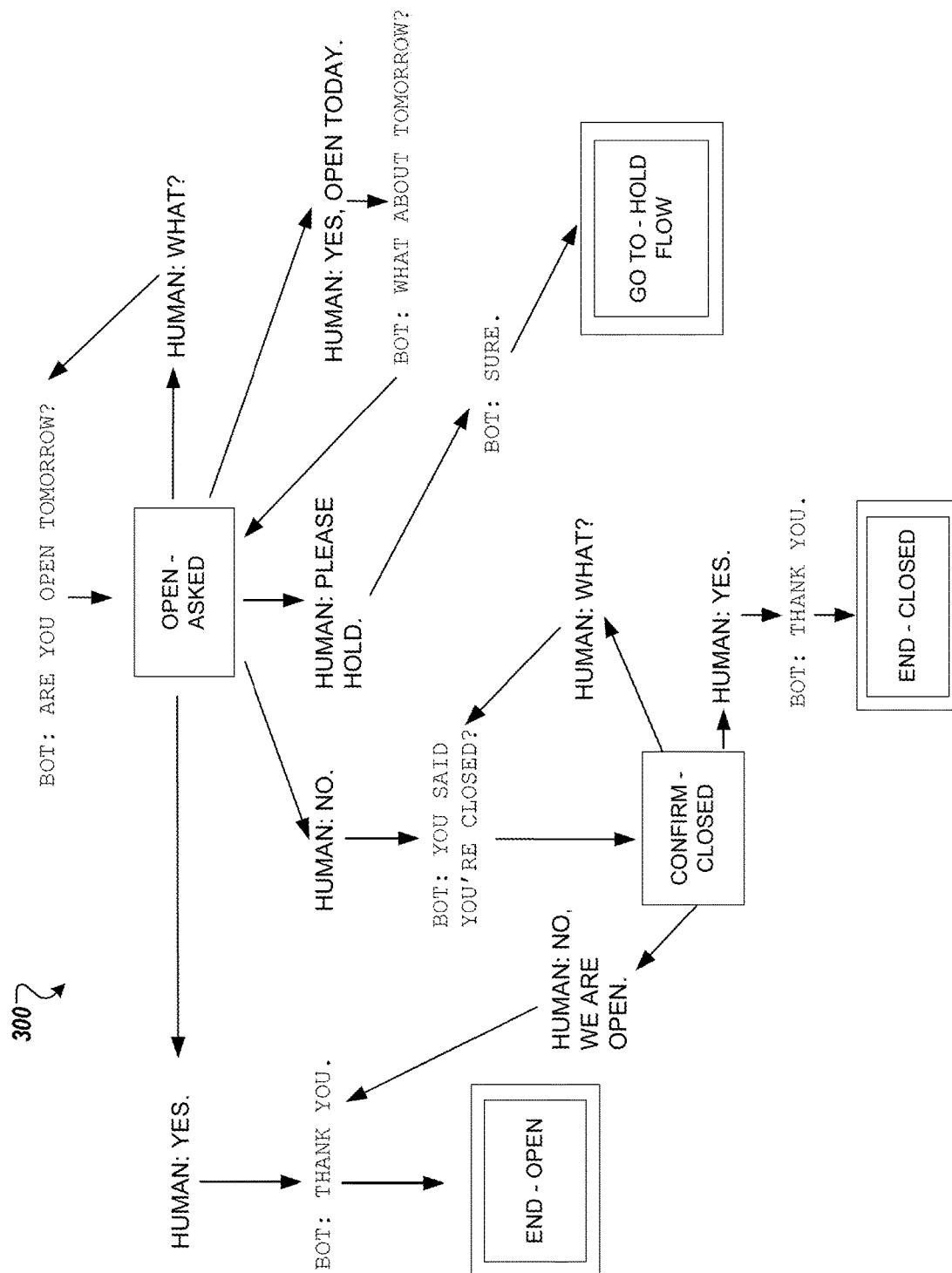


FIG. 3

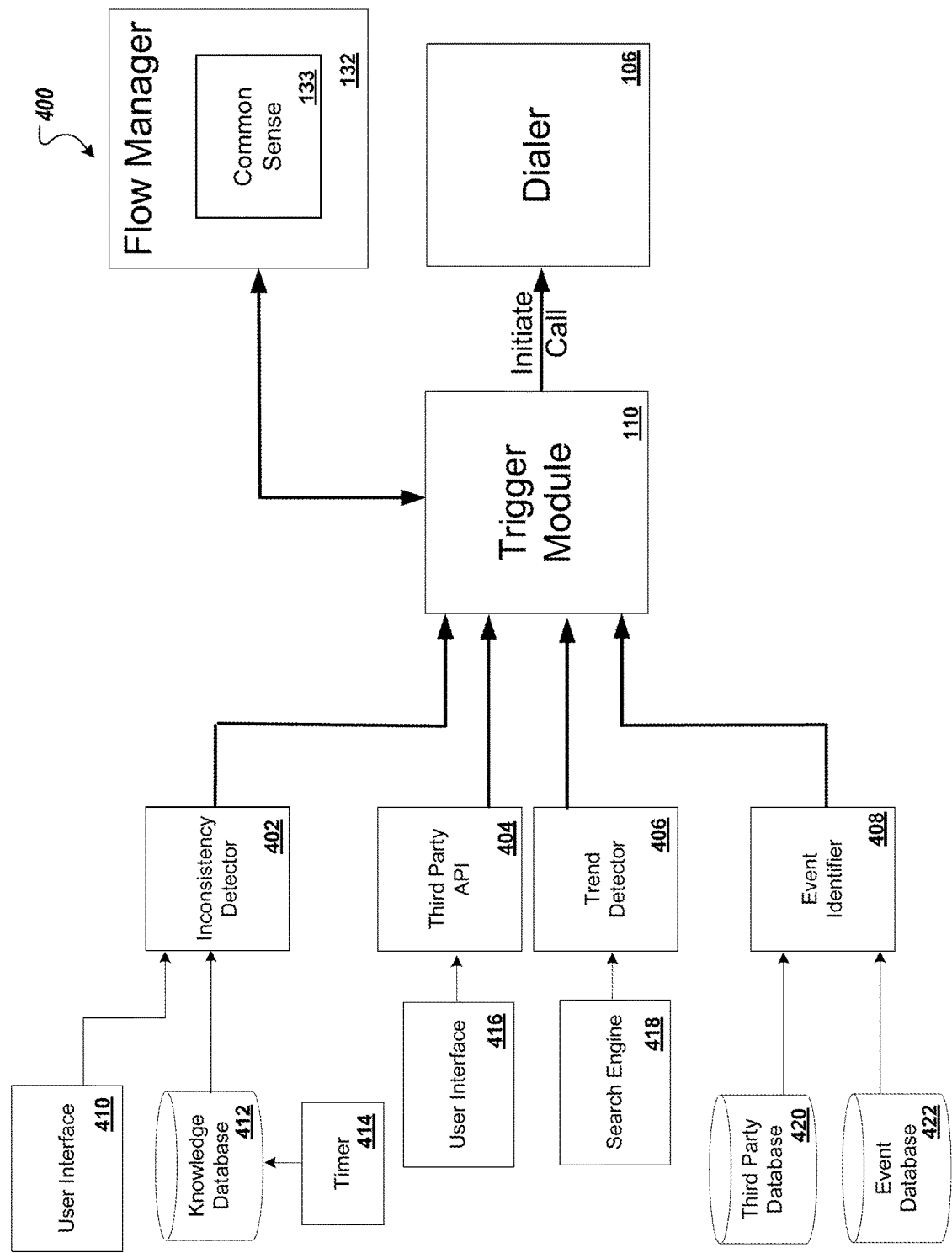


FIG. 4



500 ↗

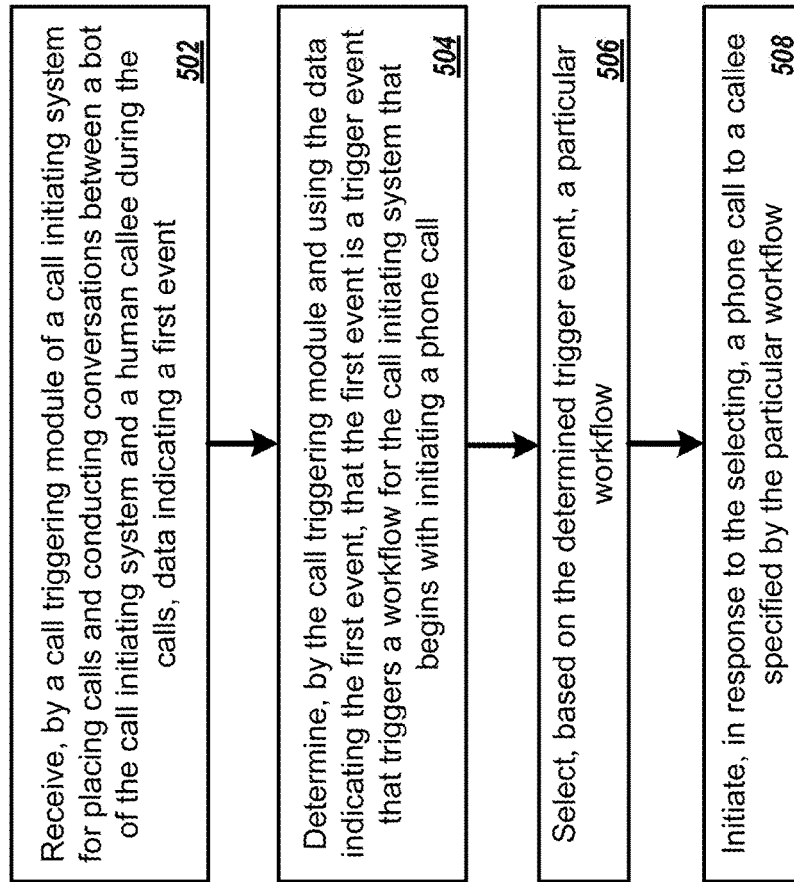


FIG. 5

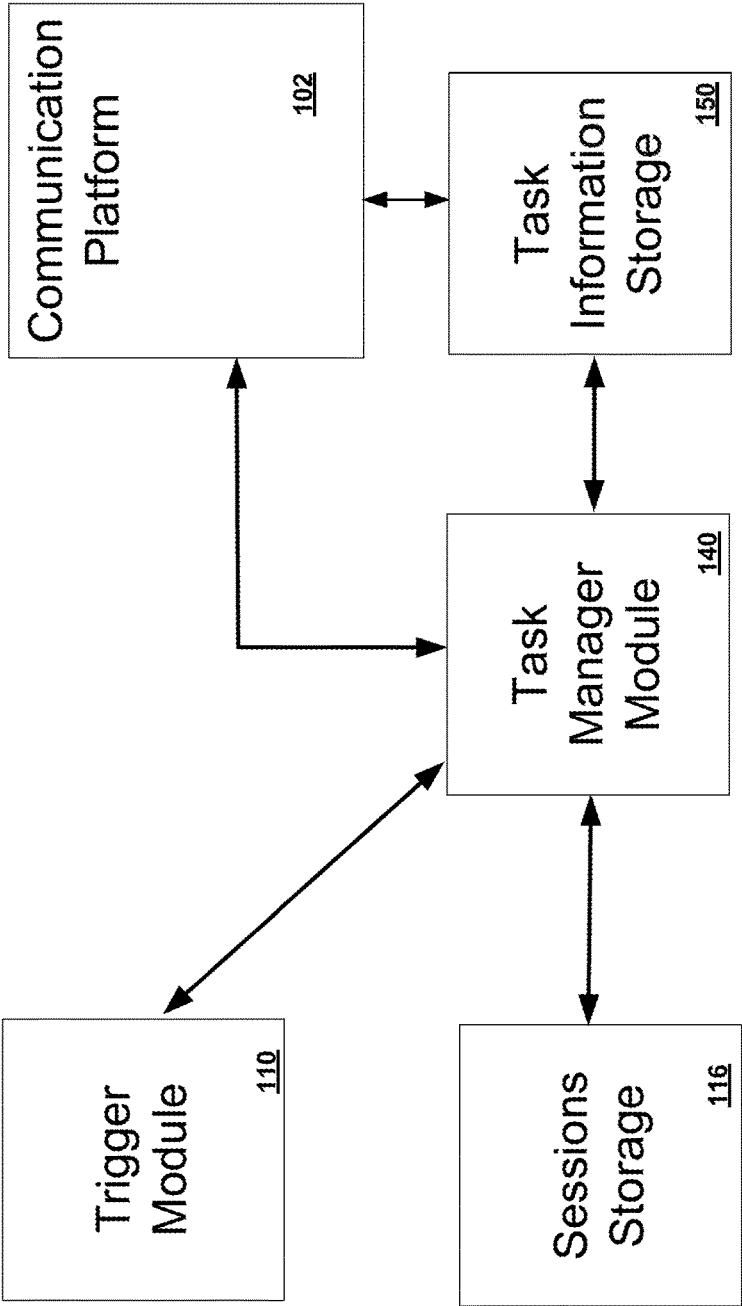


FIG. 6

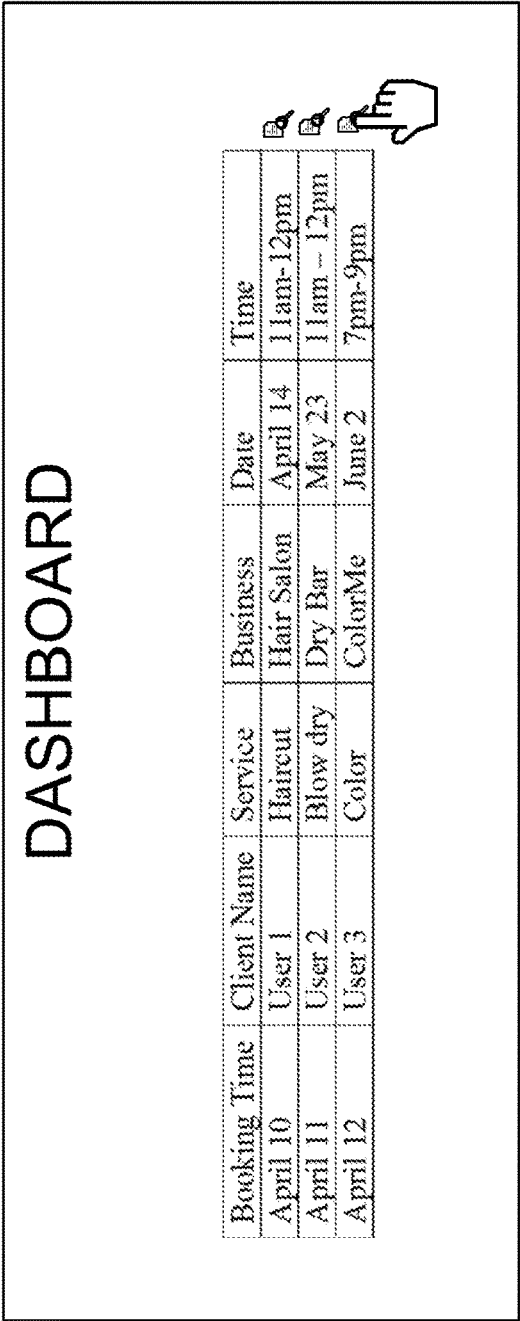


FIG. 7A

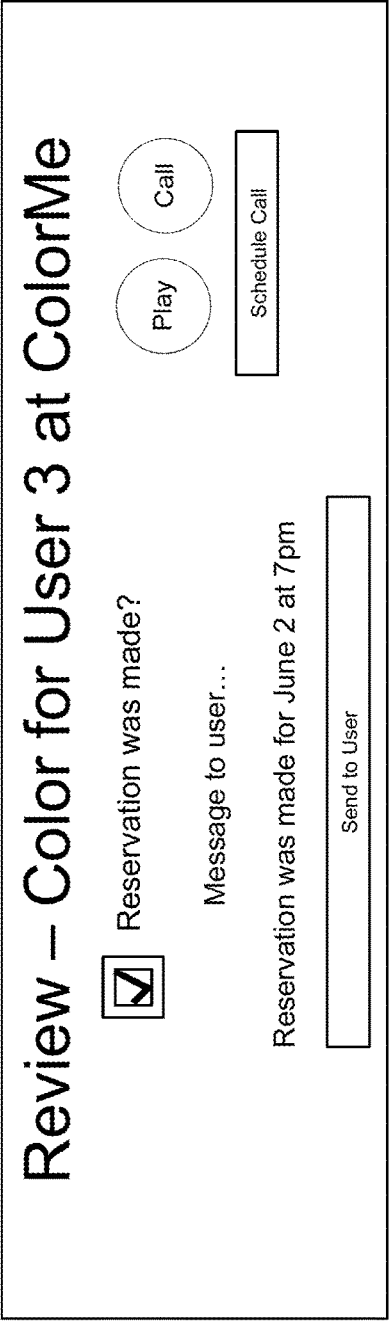


FIG. 7B

800 ↗

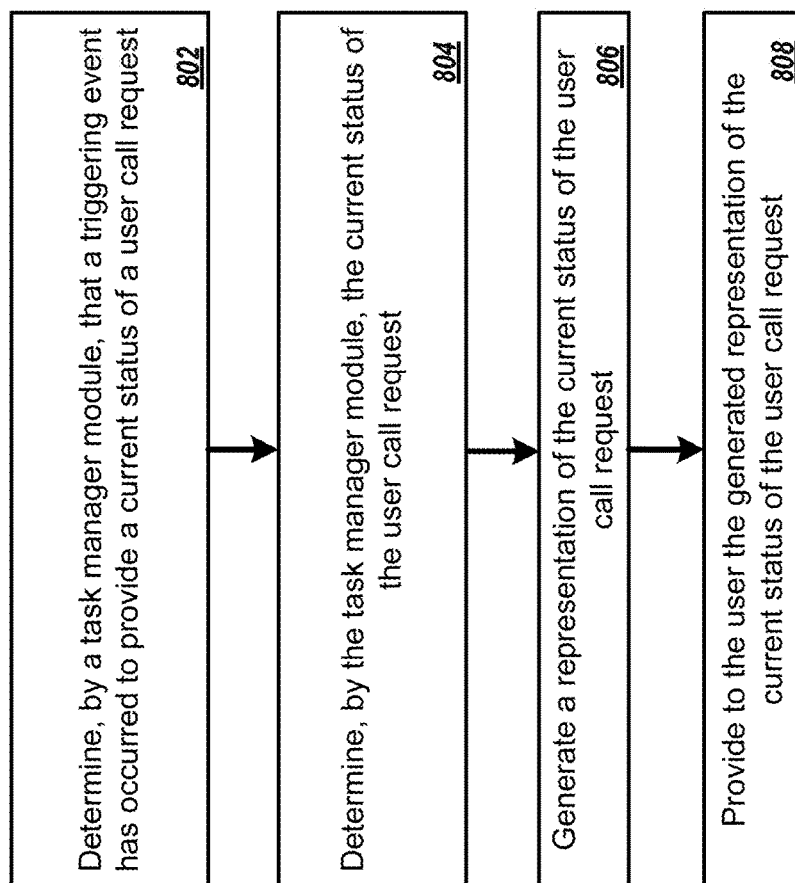


FIG. 8

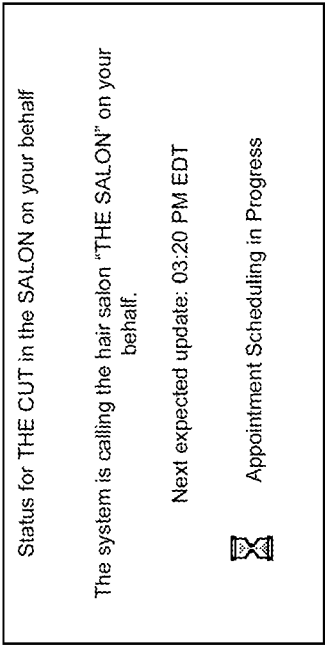


FIG. 9A

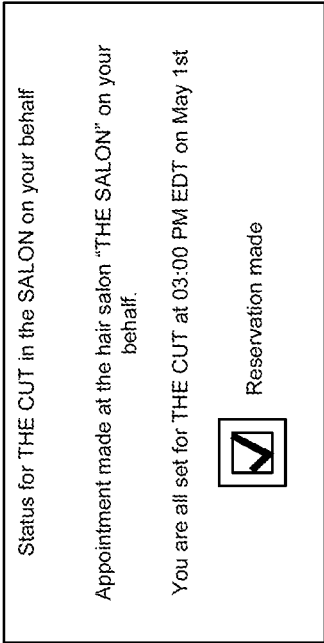


FIG. 9B

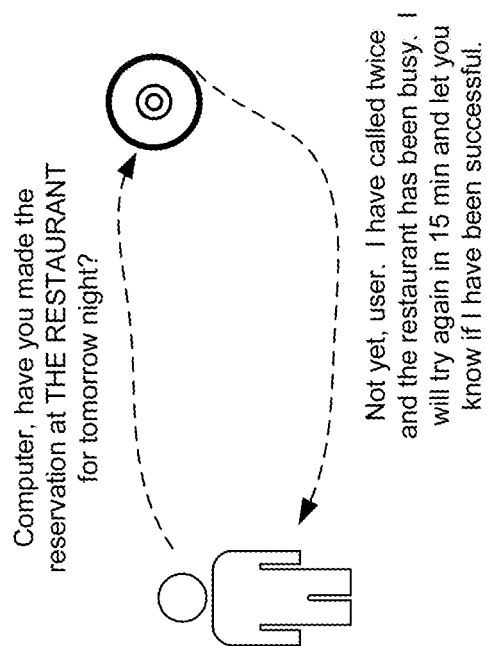


FIG. 10A

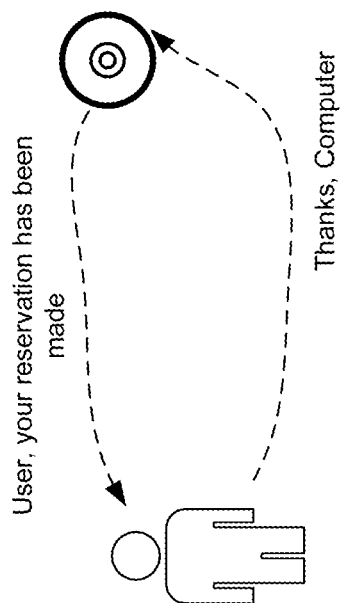


FIG. 10B

1100 ↗

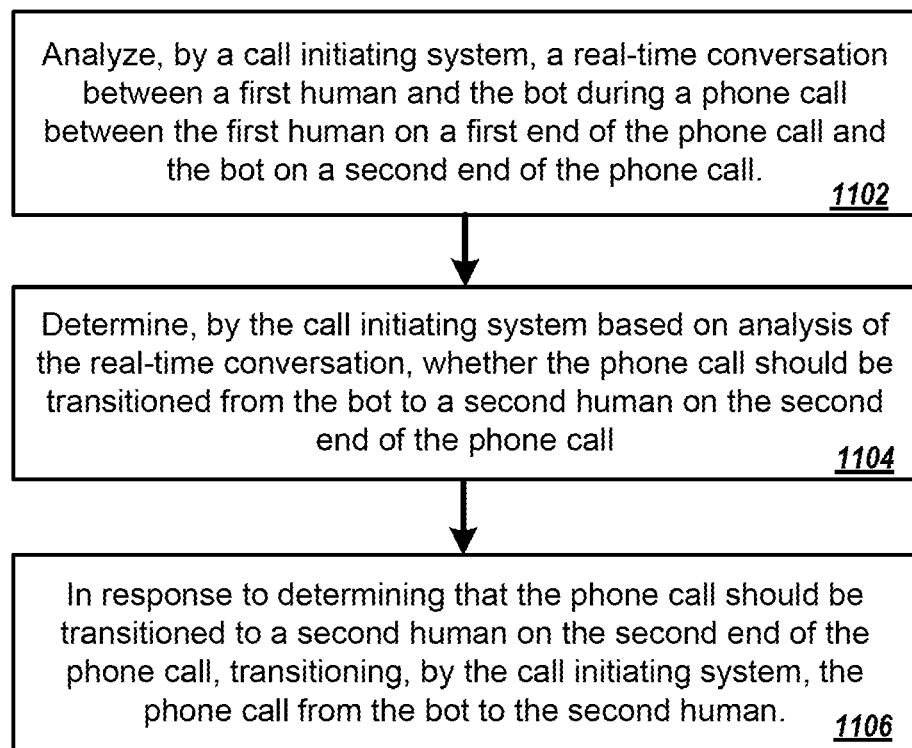


FIG. 11

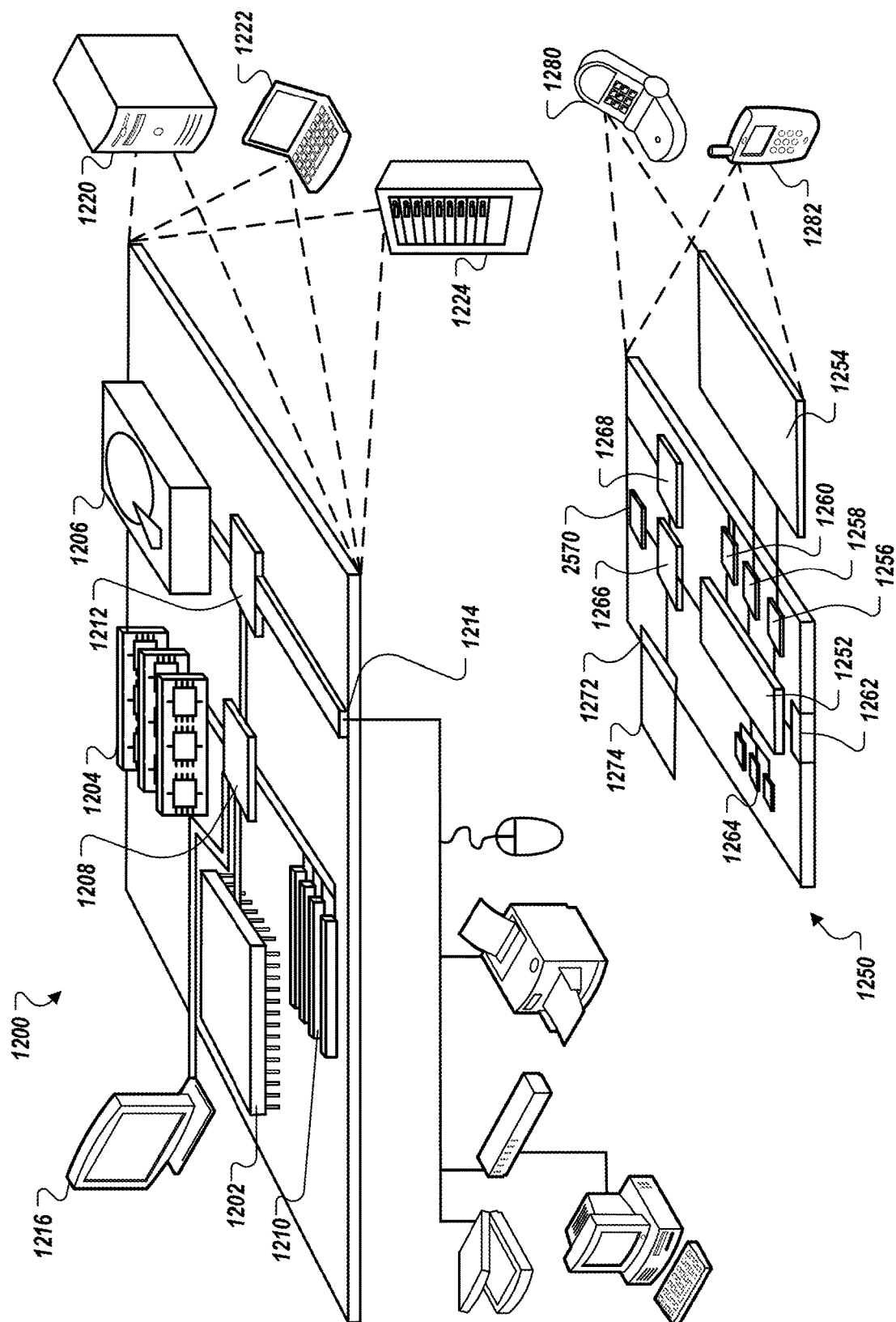


FIG. 12



## AUTOMATED CALL REQUESTS WITH STATUS UPDATES

### FIELD

**[0001]** This specification relates to natural language processing.

### BACKGROUND

**[0002]** Users may need to collect types of information that is not easily obtained without human interaction. For example, in order to verify or collect data from multiple places of business or organizations, a user may need to call each of the businesses or organizations in order to gather the information. While web search engines can assist users with such tasks by providing contact information for a service or business, the user must still call the service or business themselves to complete the task themselves.

**[0003]** In order to maintain a database of information gathered from multiple places of business or organizations, a human operator can initiate automated calls to large numbers of businesses to collect data, but selecting the callees (e.g., all restaurants in a particular town that serve the same cuisine) and placing the calls can be time-consuming when performed manually. Moreover, determining when and whether to place the calls generally requires human analysis of existing data in order to identify a need for verification, updating, or supplemental information.

**[0004]** Users may also wish to perform tasks such as make appointments or hire a service. However, there is generally a person with whom a user must interact to complete the desired task. For example, a user may be required to call and speak with a hostess in order to make a reservation at a small restaurant that does not have a website. In some cases, even when users place the calls themselves, they may encounter automated phone trees that often accept only a limited set of user responses.

### SUMMARY

**[0005]** A system can assist a user with various tasks that involve communicating with a human through a phone call, or with an automated system operated through a phone (e.g., IVR), by determining, from data received by the system, whether to initiate a call to a particular number. Once a call is placed, the system may obtain information, provide information to a third party, execute an action, e.g., on behalf of the user, and so forth. In certain examples, the system participates in a dialog with a human on the user's behalf. The dialog can occur via a telephone connection between the system and the human. In certain examples, the system may include, operate with, or form a part of a search engine following a workflow associated with intents of search engine users who submit queries involving tasks to be completed. The system may execute the tasks for the users through at least one autonomous or semi-autonomous software agent ("bot") operations.

**[0006]** In one general aspect, a method includes receiving, by a call triggering module of a call initiating system for placing calls and conducting conversations between a bot of the call initiating system and a human callee during the calls, data indicating a first event; determining, by the call triggering module and using the data indicating the first event, that the first event is a specific trigger event of a plurality of possible trigger events that triggers a workflow for the call

initiating system that begins with initiating a phone call; selecting, based on the determined trigger event, a particular workflow from multiple possible workflows, the particular workflow corresponding to the determined trigger event; and in response to the selecting, i) initiating a phone call to a callee specified by the particular workflow, and ii) executing the workflow as a two-way conversation between the bot and the callee.

**[0007]** Implementations may include one or more of the following features. For example, the determined trigger event is an inconsistency of a value associated with a first data source and a corresponding value associated with a second data source. The data indicating the first event may be provided by a user. The determined trigger event may be a user request. The determined trigger event may be a particular type of event that is one of: a weather event, an entertainment event, or a seasonal event. The determined trigger event may be a trend detected in search requests submitted to a search engine. The determined trigger event may be a passing of a predetermined period of time.

**[0008]** In another general aspect, a method includes determining, by a task manager module, that a triggering event has occurred to provide a current status of a user call request; determining, by the task manager module, the current status of the user call request; generating a representation of the current status of the user call request; and providing, to the user, the generated representation of the current status of the user call request.

**[0009]** Implementations may include one or more of the following features. For example, the determined trigger event may be a user request for status. The determined trigger event may be an operator interaction to provide status to a user after the operator has reviewed session information associated with the user call request. The determined trigger event may be a status update event. The representation of the current status may be a visual representation. The representation of the current status may be an oral representation. Providing the user the generated representation of the current status of the user call request may include determining a convenient time and method for delivering the current status to the user.

**[0010]** In another general aspect, a method for transitioning a phone call away from a bot includes: analyzing, by a call initiating system, a real-time conversation between a first human and the bot during a phone call between the first human on a first end of the phone call and the bot on a second end of the phone call; determining, by the call initiating system based on analysis of the real-time conversation, whether the phone call should be transitioned from the bot to a second human on the second end of the phone call; and in response to determining that the phone call should be transitioned to a second human on the second end of the phone call, transitioning, by the call initiating system, the phone call from the bot to the second human.

**[0011]** Implementations may include one or more of the following features. For example, analyzing the real-time conversation between the first human and the bot during the phone call may comprise: determining strain during the phone call based on conduct, demeanor, tone, annoyance level, language, or word choice of the first human. The method may include determining an increase in strain during the phone call when the bot repeats itself, apologizes, or asks for clarification. The method may include determining an increase in strain during the phone call when the human

corrects the bot or complains about the quality of the call. The method may include determining a decrease in strain during the phone call when the bot appropriately responds to dialog of the first human. Analyzing the real-time conversation between the first human and the bot during the phone call may include determining a confidence level of the call initiating system that a task of the phone call will be completed by the bot. Analyzing the real-time conversation between the first human and the bot during the phone call may include determining that the first human asked for the phone call to be transitioned to another human. Analyzing the real-time conversation between the first human and the bot during the phone call may include determining that the first human has mocked the bot or asked if the bot is a robot. Determining whether the phone call should be transitioned from the bot to a second human may include determining that the strain is above a predefined threshold; and in response to the strain being above a predefined threshold, determining that the phone call should be transitioned from the bot to the second human. Analyzing the real-time conversation between the first human and the bot during the phone call may include keeping track of one or more events in the conversation. Determining whether the phone call should be transitioned from the bot to a second human may include using a feature-based rule set that determines whether the one or more events in the conversation meet criteria of a rule; and in response to determining that the one or more events in the conversation meet the criteria of a rule, determining that the phone call should be transitioned from the bot to the second human.

**[0012]** Analyzing the real-time conversation between the first human and the bot during the phone call may include identifying intents from the conversation and identifying historical intents and historical outcomes from previous conversations. Determining whether the phone call should be transitioned from the bot to a second human may include sending intents from the conversation, historical intents, or historical outcomes to one or more machine learning models; and determining whether the phone call should be transitioned based on the intents, historical intents, or historical outcomes. The second human may be a human operator. The bot may use the same voice as the human operator so that the transition from the bot to the second human is transparent to the first human. The second human may be the user for which the bot is conducting the phone call. The method may include terminating the phone call when transitioning the phone call from the bot to the second human takes longer than a predetermined amount of time. The method may include terminating the phone call instead of transitioning the phone call to a human.

**[0013]** Other implementations of this and other aspects include corresponding methods, apparatus, and computer programs, configured to perform the actions of the methods, encoded on computer storage devices. One or more computer programs can be so configured by virtue of having instructions that, when executed by data processing apparatus, cause the apparatus to perform the actions.

**[0014]** Particular embodiments of the subject matter described in this document can be implemented so as to realize one or more of the following advantages. The amount of data storage required for various data sources is reduced because only one set of confirmed data is stored instead of multiple sets of unconfirmed data. For example, instead of storing three different unconfirmed sets of business hours for

a particular grocery store (e.g., one set collected from the storefront, one set collected from the store's website, and one set collected from the store's answering machine), a data source can store one set of confirmed store hours obtained from a call to a human representative of the grocery store.

**[0015]** By automatically detecting trigger events that indicate to the call initiating system that a call is to be initiated, the amount of human input required to perform operations such as collecting data from callees, scheduling appointments, or providing information to third parties is reduced. Additionally, because calls are only initiated when a trigger event occurs, the amount of computer resources needed to maintain a database of information is reduced due to the reduction in calls placed. The system automatically places calls to particular callees or sets of callees, reducing the amount of analysis that a human must perform and the amount of data that a human must monitor.

**[0016]** Additionally, the system conducts conversations on behalf of human users, further reducing the amount of human input required to perform particular tasks. The call initiating system can coordinate multiple calls at the same time. For example, a user may wish to make a reservation for thirty minutes in the future. The system can call each restaurant specified by the user and carry out a conversation with a representative on the other line. An employee of a first restaurant called may suggest that a reservation can be made, but the diner must sit at the bar. An employee of a second restaurant called may suggest there is a wait time of twenty minutes, and an employee of a third restaurant called may inform the system that the third restaurant requires that diners finish their meals within one hour, and therefore a table will be ready within the hour. The system can make the call to each of the three restaurants in parallel, consult the user by presenting his options and receiving a response, and make a reservation at the restaurant most suitable for the user based on his response while declining all other reservations. The automated call initiating system is efficient than a human counterpart since the automated system can make these calls all at once. The human assistant is not able to easily make all of these restaurant calls in parallel.

**[0017]** The details of one or more embodiments of the subject matter described in this specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0018]** FIG. 1A shows an example block diagram of a system for a call initiating system that places calls and conducts conversations between a bot of the call initiating system and a human during the calls.

**[0019]** FIG. 1B shows an example block diagram of a system for a call initiating system that places calls and conducts conversations between a bot of the call initiating system and a human during the calls.

**[0020]** FIG. 1C shows an example user interface through which a user may enter more details about the request.

**[0021]** FIG. 1D shows an example of a user speaking to a bot in order to make a request.

**[0022]** FIG. 2A is a flow diagram that shows an example of a process for completing a task assigned by a user.

[0023] FIG. 2B is a flow diagram that shows another example of a process for completing a task assigned by a user.

[0024] FIG. 3 shows an example workflow for a process executed by the system.

[0025] FIG. 4 is a block diagram of the triggering module.

[0026] FIG. 5 is a flow diagram that shows an example of a process for initiating a phone call.

[0027] FIG. 6 is a block diagram of the task manager module of the system.

[0028] FIG. 7A illustrates an operator dashboard that shows information about progress of existing tasks.

[0029] FIG. 7B illustrates an operator review screen for reviewing one of the user requested tasks.

[0030] FIG. 8 is a flow diagram that shows an example of a process for providing the status of a task.

[0031] FIG. 9A shows a visual status of the haircut appointment request of FIG. 1B while the appointment scheduling is in progress.

[0032] FIG. 9B shows a visual status of the haircut appointment request of FIG. 1B once the appointment has been successfully scheduled.

[0033] FIG. 10A illustrates an oral status request and update of the restaurant reservation request of FIG. 1C.

[0034] FIG. 10B illustrates an oral status update provided by the system without prompting by the user for the restaurant reservation request of FIG. 1C.

[0035] FIG. 11 shows an example process 1100 for transitioning a phone call from a bot to a human.

[0036] FIG. 12 is a schematic diagram that shows an example of a computing device and a mobile computing device.

[0037] Like reference numbers and designations in the various drawings indicate like elements.

#### DETAILED DESCRIPTION

[0038] The present disclosure describes a technology that allows an automated or semi-automated system, referred to herein as a “bot,” to communicate with people, by placing calls and independently conducting conversations with a human during the calls. The bot receives and monitors data to detect trigger events that indicate that a call should be initiated. The bot works through predefined workflows, or sequences of repeatable patterns of operations, each linked by abstract descriptions of operations to be performed, or intents. Essentially, the bot is able to use these workflows to determine how to react and what to say to a human in order to perform helpful tasks for a user.

[0039] The system handles various tasks, received as queries, such as: “book a table for two at Yves Saint Thomas Restaurant for Thursday,” “my sink is leaking and I need a plumber! It’s after 10 p.m.!” , etc.

[0040] A user who wishes to schedule an appointment, purchase an item, request services, etc., may be required to perform multiple searches and place many calls before completing the task they set out to accomplish. In a first use case, booking a table at a restaurant, the user might search for the restaurant on a search engine. In some examples, if the restaurant happens to be on a website or application, the query may be executed on the website or application (or through integration with the website or application), and if not, the user might call the restaurant and negotiate the reservation.

[0041] As an example, the system may be used to place calls for a user. The system communicates with businesses and other services to complete tasks requested by users. In some examples, the bots perform much of the communication. In some examples, a human operator may review and verify the success of operations performed by the bots. In some examples, the human operator performs the actions, and the bots learn from the human operators’ communications to improve their automated communication skills.

[0042] In a second use case, a user wishes to find a plumber outside of normal business hours. Such queries may be more difficult to process. For example, if the user were to search for a plumber manually, he might search for plumbers on a search engine and give a few of them a call. The user may have to explain to each plumber the time constraints, where they are located and the nature of the issue, and obtain a price estimate. This can be very time consuming.

[0043] Similarly, for a third use case, checking if a local store has a product in stock, the user might have to search for local stores and call each one to determine whether the store has the particular item or product they are looking for.

[0044] In addition to assisting users with specific tasks, the system can update an index of information, such as business open hours, services offered, etc. The system can be automatically triggered to update data responsive to detecting missing, aging, inconsistent data, etc. Generally, in order to obtain such information, a user may need to individually check each business or data source.

[0045] The system provides many advantages, including reducing the amount of human input required to complete certain tasks that involve initiating a phone call. For example, the system can automatically initiate phone calls based on determining that certain triggering criteria have been met, such as an inconsistency between services offered by a salon and services listed on a third party reservation website. The system can reduce friction for transaction queries, for example, by detecting frustration or discomfort of a human on one end of the phone call and ending the call or making a change to the way the conversation is conducted. The system can connect users in developing countries with services, such as transportation or educational services. The system can also connect users with low tech industries that don’t have websites or digital presences. Furthermore, the system is scalable for different applications, even compared to the largest aggregators.

[0046] FIG. 1A shows an example block diagram of a system for a call initiating system that places calls and conducts conversations between a bot of the call initiating system and a human 104 during the calls. Each component shown in diagram 100 will be described in detail below.

[0047] The system 100 includes various components and subsystems that work together to allow a bot to communicate effectively with the human 104. The system 100 may include a communication framework or platform 102, a dialer 106, a sound system 108, a call triggering module or trigger module 110, an audio package 112, a session recorder 114, a sessions storage 116, a text-to-speech module 118, a speech endpoint detector 120, stored text-to-speech results or recordings 122, an intent-to-text module 124, a speech-to-text module 126, a speech application program interface (API) 128, a text-to-intent module 130, a flow manager 132, an operator controller 134, and a bailout module 136. In some implementations, the system includes all modules. In other implementations, the system includes a combination of

these modules. For example, in one implementation, the text-to-intent layer is unnecessary and the intent is directly given to the speech synthesis module.

**[0048]** FIG. 1B shows an alternative example block diagram of a system for a call initiating system that places calls and conducts conversations between a bot of the call initiating system and a human during the calls. In this example, the communication platform **102** is replaced with a client entry point for user requests and a telephony signaling server for other requests, i.e., inbound calls from business. The system sends both types of requests to a telephony server (**196**) that makes the call with a bot service (**195**) conducting the call from the other end. In some implementations, the bot service (**195**) includes a dialog model (**198**) and a language model (**199**) to enable the bot service to conduct human-like phone conversations. The telephony server (**196**) may include a TTS model (**197**). A speech recognizer (**191**) and/or audio mixer (**194**) may provide information for the telephony server (**196**) to understand and respond back to the human on the other end of the phone call (**190**). Operators (**134**) monitor calls using a task user interface (**160**) and a curation user interface (**170**). The operators (**134**) can review recorded calls from the recording studio and evaluation TTS (**114**). A call player (**162**) will replay calls back to the operators (**134**). Operators can schedule calls using a local agent (**175**) to initiate phone calls through the telephony server (**196**).

**[0049]** In the implementation of FIG. 1A, the communication platform **102** allows the bot to contact external actors by performing tasks such as placing calls, receiving inbound calls from businesses or users (**104**, **144**), or contacting target business. The communication platform **102** also allows the bot to receive requests from a user to make calls on the user's behalf.

**[0050]** In some implementations, a user requests a call to another user or business through interaction with a user interface or through a speech request. These user requests may be for assistant-type tasks such as: booking an appointment, making a restaurant reservation, finding a dog walker, or figuring out which store has an item that the user would like to purchase.

**[0051]** FIG. 1C shows an example user interface through which a user may enter more details about the request. A user may initiate a request by clicking a "book" button or interacting with the user interface in some other way. For example, if the user would like to make a haircut appointment, the user may interact with a website associated with a salon where the user would like to get a haircut. The user may alternatively interact with a search results list that includes the salon as a result in the search results or with a user interface showing the salon on the map. Any of these interfaces may allow the user to request a call. The user may enter the details of the user's request such as: the professional stylist the user would like to see, the category for the service the user would like to have done, and the date and time for the service. As shown in FIG. 1B, the user may click a "Continue with booking" button or take some other action to indicate that a request for a hair appointment is being made.

**[0052]** FIG. 1D shows an example of a user speaking to a bot in order to make a request. After the user speaks a request to the bot, the bot may acknowledge the request. The bot may also request additional information about the request from the user. For example, if the user speaks a

request to the bot about making a haircut, the bot may ask for information regarding where the user would like the haircut appointment, the day the haircut appointment should be scheduled, and what kind of haircut service the user would like to schedule.

**[0053]** A user may make a task request to the system at a time when the task cannot be performed. For example, the user may request a call to schedule a haircut appointment at 11 pm at night when all haircut salons are closed. Therefore, the system may store the request in the task information storage to be initiated and completed at a later date or time, such as during the open hours for the salon as otherwise determined by or obtained by system **100**.

**[0054]** In some implementations, the system provides the user with initial feedback that there will be a delay in processing the request. For example, when the user makes the request to call to schedule a haircut appointment at 11 pm when the salons are closed, the system provides the user with a visual, audio, or some other indication that since the salon is closed, there will be a delay in completing the task until the system can reach the salon once it opens.

**[0055]** In some implementations, the task information storage **150** stores information about each task such as: the name of the user requesting the task, the one or more people or places to call, the type of task requested, the method in which the task request was made, details about the task that are type specific, details about the activities that have been done to complete the task, the start date of the task, the completion date of the task, the time of the last status update to the requesting user, the operator who double-checked the call task, the user requested end date for the task, and the current status of the task.

**[0056]** In some implementations, a task manager module **160** determines when to schedule calls to people or businesses. The task manager module **160** monitors tasks from the task information storage **150** and determines the appropriate time to schedule the received tasks. Some tasks are immediately scheduled while other tasks are scheduled after certain triggering events occur.

**[0057]** In many situations, there will be a human, such as the human **104**, on the other end of a call placed by the system **100**. The human **104** can be a representative of an organization that the bot is attempting to contact. In some examples, communication platforms are used in order to call businesses. The present system **100** can be integrated with the communication platforms. For example, the present system **100** can use a framework for testing web applications to programmatically operate a web browser and use a web-based teleconferencing service. The system **100** can create and use several communication platform accounts. In some examples, the system **100** can automatically alternate between different communication platform accounts to avoid throttling of call speed.

**[0058]** The dialer **106** facilitates initiating, or placing, the calls that the bot conducts. The dialer **106** is communicably connected to the communications platform **102**. The dialer **106** provides instructions to the communications platform **102** to initiate a phone call to a particular callee selected by the dialer **106**. For example, the dialer **106** can play audio tones corresponding to digits of a phone number. Once a call is placed, the system **100** can conduct a conversation with the human callee on the other end of the line.

**[0059]** The dialer **106** can receive instructions to initiate a call to a particular callee. For example, the dialer **106** can

receive data containing instructions from other modules in the system **100** such as the trigger module **110** or the flow manager **132**.

**[0060]** The trigger module **110** detects trigger events, or particular events that indicate that the system **100** should initiate a call to a particular callee. The trigger events can be events of a predetermined type. For example, a user of the system **100** can specify particular types of trigger events. The trigger events can include explicit actions performed by a user of the system **100**, a detected pattern in data provided to the trigger module **110**, a predetermined period of time passing since a particular event occurred, and various other types of events. In response to detecting a trigger event, the trigger module **110** provides instructions to the dialer **106** to initiate a call to a particular callee, or to the flow manager **132** to select a node of a particular workflow or provide instructions to the dialer **106**.

**[0061]** The sound system **108** is used to record and play audio. In some examples, three virtual streams are set up: (a) incoming audio from a phone or teleconferencing service to the system **100**; (b) outgoing audio from the system **100** back to the communication platform; (c) a mixed stream combining a and b, and are used to record the entire call. The sound system **108** uses the audio packages **112** to perform the communications through the communications platform **102**.

**[0062]** The audio packages **112** are used to communicate with the sound system **108**. In some examples, the present system **100** includes an audio module that wraps the audio package **112** and handles a continuous stream of incoming audio packets. The module also records every incoming packet, and allows to playback pre-recorded audio files. The present system **100** uses various bit depths, sampling frequencies, packet sizes, etc.

**[0063]** The system **100** can record incoming and outgoing conversations conducted by the bot. The audio package **112** can enable the system **100** to record a particular session, or call, using the session recorder **114**. In some examples, the session recorder **114** can record the portion of the conversation conducted by the bot by recording the bot's speech as it is generated. In other examples, the session recorder **114** can record the portion of the conversation conducted by the bot by recording the bot's speech externally as it is output by the communications system **102** to the human **104**. The session recorder **114** can record responses of the human **104** as well.

**[0064]** The session recorder **114** stores the recorded session data in the sessions storage **116**. The recorded session data can be stored as audio data or as feature data that represents the audio data. For example, the recorded session data can be stored as vectors that store values for particular features of the audio data of the session. The sessions storage **116** can be a local database, a remote server, physical memory within the system **100**, or any of various other types of memory.

**[0065]** A speech endpoint detector **120** simplifies conversations between the bot and a human on the other side of the line. In order to simplify the conversation, it is sectioned into individual sentences, discretely switching between the human and the bot. The speech endpoint detector **120** is responsible for receiving the continuous input audio stream from the audio package **112** and converting it into discrete sentences.

**[0066]** The speech endpoint detector **120** detects end-points of speech. In one implementation, the speech endpoint detector **120** operates in two states: waiting for speech; and waiting for silence. The speech endpoint detector **120** alternates between these states as follows: each audio packet is examined by comparing its root-mean-square-deviation (RMSD) to a predefined threshold. A single packet is considered as "silence" if its RMSD is below this threshold. Whenever a non-silence packet is received, the module will switch from the "waiting for speech" state to the "waiting for silence" state. The module will switch back only after a period of consecutive silence packets lasting a predefined period of time, depending on the state of the entire system, is received.

**[0067]** In some implementations, during the "waiting for sound" period, the speech endpoint detector **120** fabricates pure-silence packets (one fabricated packet per ten real packets) and transmits them to the speech-to-text module **126**. The fabricated packets can avoid disconnections from the speech API **128**. During the "waiting for silence" period, the speech endpoint detector **120** transmits packets of silence from the stream for up to a predefined period of time (useful for baseline noise estimation) and then transmits all of the audio packets.

**[0068]** In other implementations, the speech endpoint detector **120** uses machine learning, neural networks, or some form of deep learning trained to observe the intonation and language context to find endpoints.

**[0069]** In some examples, the speech endpoint detector **120** considers what was said, the intonation of the speaker, etc., when determining how to parse a particular stream of audio input. For example, the speech endpoint detector **120** can determine that a particular human callee **104** has a tendency to end sentences with a low inflection, and speech endpoint detector **120** can predict the end of a sentence spoken by the callee **104** when a drop in inflection is detected. A speech endpoint detector **120** can adjust thresholds dynamically during the call based on the signal-to-noise ratio in a time frame.

**[0070]** The speech-to-text module **126** converts the audio data parsed by the speech endpoint detector **120** into text that can be analyzed for intent that is used to select the bot's next response. The output of the speech-to-text module **126** is an ordered list of speech options, and in some cases, a confidence for the best option is provided. The speech recognition process includes two major components: the acoustic module and the language module. For the acoustic module, the system can use a model trained from recordings of people talking directly to their phone. A neural network may be used by the model, and in some examples, the first layer of the neural network may be re-trained to account for the vocoders present in a phone call. A vocoder is a voice codec that produces sounds from an analysis of speech input. The neural network may also be re-trained to account for background noise which differs between calls to businesses and personal phone calls. The language module may be built using a system that biases the language module based on past experience of the system. In some examples, the bias may be configured automatically. In some examples, this bias is configured manually. In some examples, the language-bias configuration changes between verticals.

**[0071]** The speech-to-text module **126** uses the context of a call in order to bias the language module based on what the person from other side of the conversation is predicted to

say. For example, the system's bot asks, "Are you open on Tuesday?" Based on this question, there is a high likelihood that the person on the other side of the conversation will respond with an answer such as, "no, we are closed" or "yeah, sure." The bot learns the likely responses based on past calls and uses predictions to understand the incoming audio. The bot can predict full sentence responses, but the bot may also predict phrases. For example, after the bot says: "we have seven people in our party," the bot may expect the phrase: "you said seven people?" The bot may also expect the phrase, "you said eleven?" since seven and eleven sound similar. The bot may also expect with some lower chance for a response such as, "you said two?" The bot can assign probability weight for each phrase based on its predictions.

**[0072]** In some implementations, the speech-to-text module 126 uses the speech API 128 to convert the audio data into text. In some examples, the speech API 128 uses machine learning to convert the audio data into text. For example, the speech API 128 can use models that accept audio data as inputs. The speech API 128 may use any of a variety of models such as decision trees, linear regression models, logistic regression models, neural networks, classifiers, support vector machines, inductive logic programming, ensembles of models (e.g., using techniques such as bagging, boosting, random forests, etc.), genetic algorithms, Bayesian networks, etc., and can be trained using a variety of approaches, such as deep learning, perceptrons, association rules, inductive logic, clustering, maximum entropy classification, learning classification, etc. In some examples, the speech API 128 may use supervised learning. In some examples, the speech API 128 uses unsupervised learning. In some examples, the speech API 128 can be accessed by the speech-to-text module 126 over a network. For example, the speech API 128 can be provided by a remote third party on a cloud server.

**[0073]** To address the synchronizing of dialog, such as determining the context in which a person was speaking in order to determine natural opportunities for response by the bot, the system 100 can identify an intent. An intent is a formal-language representation of a single semantic meaning in a sentence, either said by a human or by the bot. In some implementations, the system 100 ignores any intent received from a human between the last intent received and the bot's reply in order for the bot to produce a response relevant to the most recent sentence spoken by the human. The system 100 can, however, use previous intents to inform future responses. For example, the system can mark intents received before the most recent intent received as ANCIENT, parse the ANCIENT intent, and store it for offline evaluation. In some examples, various other forms of handling logic may be used.

**[0074]** While most of the system 100 is use case agnostic, some parts of the system 100 are either manually configured or fully programmed for a specific use case, namely a system vertical. A vertical is essentially composed of a schema of intents and business logic code. Intents in the schema are an internal formal-language representation of a single semantic meaning in a sentence, either said by the human or by the bot. For example, in the opening-hours extraction vertical, we have an "AreYouOpen {date:tomorrow}" bot intent, and a corresponding "WeAreClosed {date\_range: Sep}" human intent. The process in which incoming audio from the human is transformed into an intent is referred to herein as Intent Resolution. The opposite process (transforming a bot intent

into speech) is referred to as Intent to Speech. While the schema is configured per vertical, most of the code that learns and classifies the intents is generic and used across verticals. In some examples, only the language-specific parts of the system reside in the intent resolution and intent to speech configurations.

**[0075]** The logic code can be programmed per vertical (sharing some common code) and determines the bot behavior for every possible situation, defined by the context of the call (the input parameters and what happened so far), as well as the incoming human intent. In one implementation, speech is changed to text, which is then interpreted as human intent. The human intent is used to determine robot intent. In some verticals, the bot leads the conversation, while in other cases it mostly reacts to the human. For example: in data acquisition type of verticals, the bot aims to extract some information from the business. Usually, it will try to ask a sequence of questions until it gets all desired information. For transaction type verticals, where the bot aims to, for example, make a reservation, it will predominantly answer questions originated by the human ("what is your name?" . . . "and phone number?" . . .). In such cases, the system will take the lead only if the human suddenly becomes silent, etc. Programmers can design the flow between human intents and robot intents so that the translation makes logical sense. In some implementations, there is a protocol for the flow that non-engineers would be able to control in order to change or update the translation from a human intent to a robot intent. The flow may also be learned automatically using machine learning.

**[0076]** In another implementation, the input human intent resolution can be a hidden layer and machine learning can be used to learn the output robot intents directly from the input text. The human speech input can be changed to text and then robot intents can be determined directly from this text. In still another implementation, the system can output intent directly from human speech. Both of these designs use machine learning to learn the robot intents corresponding to the context and respective inputs.

**[0077]** The text-to-intent module 130 is configured using a schema of the possible incoming intents, example sentences for each such intent, and a language-bias configuration. Essentially, the text-to-intent module 130 is responsible for "snapping" an incoming sentence into a pre-defined (or an "unknown") list of intents, while accounting for unfamiliar phrasings and for errors in the speech recognition process. For example, in some implementations, the text-to-intent module 130 can identify that the sentence (as received from the speech recognition module) "who be open eleven o'clock in the morning till nine o'clock tomorrow nine thirty I'm sorry" is similar to the known example "we open at +(TIME,from\_time) and close at +(TIME,to\_time) I'm sorry", which is an example of the intent "WeAreOpen {from\_time=11 am, to\_time=9:30}". Fields like the "from\_time" and "to\_time" are the intent parameters.

**[0078]** The text-to-intent module 130 may be composed of two main parts: (1) annotators and (2) annotated-text to intent classifier. In some implementations, the system has a post-classification phase which does argument classification. For example, for the phrase "Monday to Tuesday, sorry Wednesday, we are closed," the annotation part would rewrite the text into: "<DATE: Monday> to <DATE: Tuesday>, sorry <DATE: Wednesday> we are closed." This example shows that the phrase is rewritten with annotators

that specify the annotations in response text. The annotated-text intent classification would turn the annotated phrase into: WeAreClosed {day1: monday, day2: tuesday, day3: wednesday}. The post-classification phase would rewrite the phrase into: WeAreClosed {from\_day: monday, to\_day: wednesday, wrong\_to\_day: tuesday}.

**[0079]** As soon as the system **100** receives speech options, it annotates each one of them for dates, times, common names, etc., using the text-to-intent module **130**. This is done for two purposes: (1) extracting the intent parameters for the logic module (e.g., “time: 10 am”); (2) generalizing the text to simplify finding a match to sentences previously encountered. The text-to-intent module **130** receives output from the speech-to-text module **126** and annotates the list of speech options. The text-to-intent module **130** then uses the annotations to map the most likely option to an intent that is used by the flow manager **132** to select a next action within a particular work flow.

**[0080]** In order to reduce computation time during the call, the system **100** can build, in advance, a library of known texts that should be annotated (relative to the current time and date). For example, on Tuesday Sep. 1, 2015, “tomorrow”, “this Wednesday”, “September 2nd”, etc., may be stored as candidates for the “DATE: (2015, 9, 2)” annotation. In real-time, the system **100** iterates sequentially on the words in the input sentence (after some canonization) and search for the longest match of an annotation candidate. The system **100** then replaces the text with the annotation and return an edited string where all candidates, left to right, are replaced with annotations. For example, “we open at 7 in the morning” will be replaced with “we open at @(TIME, 7 am)”.

**[0081]** The annotator or text-to-intent module **130** can also be responsible for contractions. For example, the system **100** can encounter a sentence like “let me see . . . yes, 4, 4 pm”. The text-to-intent module **130** may replace “4, 4 pm” with a single “@(TIME, 4 pm)” annotation. Furthermore, the text-to-intent module **130** might contract small time corrections such as, “we close at 10, ah, 10:30 pm” to “we close at @(TIME, 10:30 pm)”.

**[0082]** In other implementations, the system may use other methods for annotating the text such as: a machine learning algorithm that can learn how to annotate the text based on curated data; prefix trees that can be used to annotate the text; or rule-based patterns that may be derived specifically for annotations.

**[0083]** The text-to-intent module **130** parses and annotates new sentences on most calls, and the speech recognition often distorts many of the spoken words. The system **100** has thousands of intents stored per use case, and classifies each sentence as having an intent from the stored intents that is determined to be most relevant to the sentence based on the parameters of the sentence. For example, based on detecting words in a particular sentence that suggest a question asking for the caller’s name, the system **100** can classify the particular sentence as having the intent of asking for a name. In some implementations, the system **100** does not recognize an intent for a sentence, and can classify the sentence as having an unknown intent.

**[0084]** The text-to-intent module **130** uses machine learning algorithms to handle classification. For example, the system may use a combination of a conditional random field module with logistic regression modules. In one implementation, the classification is done on the sentence level, i.e., a

string of text is converted to a set or list of intents. In another implementation, every token in the original string is classified to an intent and intent boundaries are also classified. For example, the sentence, “On Monday, we open at seven, let me think . . . on Tuesday, we open at eight,” would be classified in the first implementation as containing the intents: GiveDailyHours+AskToWait. In the second implementation, the substring “on Monday we open at seven,” would be classified as the boundaries of a GiveDailyHours intent, the substring “let me think . . .” would be classified as another intent of type AskToWait, and the substring “on Tuesday, we open at eight,” as another intent of type GiveDailyHours.

**[0085]** In some implementations, the text-to-intent module **130** may not use machine learning algorithms, but instead uses a set of examples for each intent, and then uses 1-nearest neighbor (a pattern recognition algorithm) between each speech option and all examples, where the distance metric of the algorithm is a variation of normalized edit distance (a way of qualifying how dissimilar two strings, e.g., words, are to one another) of the words of the sentence. The distance between two individual words is more complicated, and aims to be an approximation of a phonetic distance. In some examples, semantic distance may be determined by the annotated text-to-intent module **130**.

**[0086]** In practice, the text-to-intent module **130** also can also use cross speech options signals (e.g., a number that exists only in one of the speech options is likely a bad interpretation). In some examples, the text-to-intent module **130** biases the result of the annotated-text to intent module based on the system context as a prior. Finally, the text-to-intent module **130** has some tailored extractions for nebulously-defined intents like “ComplexOpeningHours” where the system is able to identify that a complex phrasing of the opening hours had been given, but the system could not correctly extract the parameters (e.g., “. . . dinner is served until 9, you can order desserts for one more hour, and the bar is opened until 2, but we don’t accept customers after 1 . . .” etc.).

**[0087]** In some examples, the examples used for classification are automatically deduced based on curated past calls, and can also be edited manually. The generalization process may replace the text with annotations and omits suspicious curations.

**[0088]** In some examples, the human does not speak a single intent, but rather a sequence of intents. For example: “you want a haircut? what time?” The example system supports any number of intents in a given sentence. In some examples, the annotated-text to intent module determines specifically positive and negative intents as prefixes for other intents (e.g., “no sir, we will be closed at this date”=>Negative+WeAreClosed). In some examples, the annotated-text to intent module supports any chaining of intents.

**[0089]** The system includes multiple modules that perform different functions of organizational logic, including the flow manager **132**, which includes a common sense module **133**, and the bailout module **136**.

**[0090]** The flow manager **132** can include custom code per vertical that tracks each call and determines how to respond to each intent received from the human (or a long silence). However, in other implementations, the flow manager is generic across verticals. The response is a list of synthetic intents to say to the human **104** (the bot can also choose to

remain silent), and sometimes a command to end the call. The flow manager **132** is also responsible for generating the outcome of the call, including any information gathered during it. In some examples, the system **100** learns how to react to each input is based on live calls, initially made by human and later on by a ‘child’ bot. The system **100** keeps the logic as flexible as possible to account for any misunderstanding it had during the call.

**[0091]** The system **100** has multiple flows, each of which is tailored for a particular type of task, such as determining opening hours for a business or making a reservation for a salon appointment. The system **100** can maintain common libraries shared between the different flows, and extract sub-flows from the history of placed calls, allowing the system to jump-start a new vertical for each task. In some examples, the system **100** may automatically learn flows for different tasks based on the manually placed calls.

**[0092]** Humans may skip some important details when speaking without confusing their conversation partner. For example, a human might say “we open 10 to 4”. The bot needs to understand whether the business opens at 10 a.m. or 10 p.m., and similarly, whether it closes at 4 pm or 4 am. If, for example, the business is a nightclub, the bot might be expected to assume 10 p.m.-4 a.m.; if the business is a restaurant, the bot might be expected to assume 10 a.m.-4 p.m., etc.

**[0093]** The flow manager **132** includes a common sense module **133** that disambiguates intents in received speech input. In some examples, the flow manager **132** includes multiple types of common sense modules, e.g., a module that learns from statistics over some data set (e.g., the baseline local database), and a module that is manually programmed. The first type of module takes a data set of options (e.g., opening hours) and calculates the p-value for each option and sub-option (e.g., “2 am-4 am” or just “2 am?”). The second type of module uses a set of predefined rules that prevent the system from making “common sense” mistakes that might reside in the dataset. Whenever there are multiple ways to interpret some variable, the flow manager **132** can combine the two scores to determine the most probable option. In some examples, the flow manager **132** concludes no option is probable enough, and the system **100** falls back on trying to explicitly ask the human to clarify what they meant.

**[0094]** The common sense module **133** can use data from similar callees to select a most probable option. For example, if most bars in Philadelphia operate from 8 p.m. to 2 a.m., the common sense module **133** can determine that the most probable option for an ambiguous phrase: “we’re open from 10 to 2” is that the speaker meant 10 p.m. to 2 a.m. In some examples, the common sense module **133** can indicate to the flow manager **132** that further clarification is required. For example, if most post offices in Jackson, Mich. have business hours of 10 a.m. to 5 p.m., the common sense module **133** may instruct the flow manager **132** to ask for clarification if the system **100** believes that the callee responded that their hours are “2 p.m. to 6 p.m.,” which is a threshold amount different from the typical post office.

**[0095]** Sometimes, there is an accumulated strain during the call, usually due to high background noise, exceptional scenarios, a heavy accent, or just a bug in the code. Strain may also be caused by unexpected intent. For example, when calling a restaurant, the system might encounter the unexpected sentence: “so, do you want to make a presenta-

tion?” or “we don’t have a TV to show the Super Bowl, just so you know.” The system should handle intents that it has previously not encountered. In order to identify problematic conditions for either party, the bot tries to quantify the amount of stress exhibited during the call. The bailout module **136** can mimic an operator supervising the call and choose when to implement manual intervention.

**[0096]** The operator controller **134** is communicably connected to the flow manager **132**, and the operator controller **134** allows a human operator to provide instructions directly to the flow manager **132**. In some examples, once a call is transferred to a human operator to handle, the operator controller **134** puts the flow manager **132** into a holding pattern or pauses or shuts down the flow manager **132**.

**[0097]** Once the flow manager **132** selects a next node within a particular workflow based on the determined intent from the text-to-intent module **130**, the flow manager **132** provides instructions to the intent-to-text module **124**. The instructions provided by the flow manager **132** include a next intent to be communicated to the callee through the communications platform **102**. The intent-to-text module **124** also produces markup cues for the speech synthesis, e.g., defining a different emphasis or prosody on some of the words. The intent-to-text module **124** can use manually defined rules or reinforcement learning to produce new text from intent.

**[0098]** The output of the intent-to-text module **124** is text to be converted into audio data for output at the communications platform **102**. The text is converted into audio by the text-to-speech module **118**, which uses previously stored text-to-speech outputs and readings **122**. The text-to-speech module **118** can select previously stored outputs from the stored outputs/readings **122**. In some implementations the system will use a text-to-speech synthesizer during the call. For example, if a common response selected by the flow manager **132** for the bot to provide is “Great, thank you for your help!” the text-to-speech module **118** can select a previously generated text-to-speech output without having to generate the output at run-time. In some examples, the text-to-speech module **118** uses a third party API accessed through a network connection, similar to the speech API **128**.

**[0099]** As discussed above, in certain examples, a user may initiate a task for the system **100** by interacting with search (e.g., web search) results provided to the user. For example, the user may look up “book a table for two tonight at a Michelin starred restaurant.” The task manager module **140** may receive a task and store the task information in the task information storage **150**. The task manager module **140** may then determine when to schedule the task and set a triggering event. For example, if the user requests to book a table before the Michelin starred restaurant is open, the task manager module **140** may determine when the restaurant is open and set a triggering event for that time. If the task manager module **140** knows that there will be a delay in processing because of the triggering event, the task manager module **140** may alert the user to the delay by providing a visual, audio, or some other indication. In some implementations, the task manager module **140** may provide information about the time it will take to complete the task, when the task is scheduled to start, or provide more information as to why the task is delayed.

**[0100]** The trigger module **110** may detect that a particular trigger event (in this example, the opening time of the



restaurant) has occurred, and instructs the dialer **106** to place a call. In some examples, the system **100** can present the user with options to select a restaurant to call. In other examples, the system **100** can automatically place a call to a particular restaurant selected based on a set of characteristics. The user can define default preferences for placing calls for particular tasks. For example, the user can designate that the system **100** should select the closest restaurant to the user's current location to call, or that the system **100** should select the most highly rated restaurant to call.

**[0101]** In certain examples, the system **100** includes or forms a part of, or is configured to communicate with, a communications application, such as a messaging or chat app that includes a user interface through which a user provides to the system requests for assistance with a task. For example, the user may be able to text a number with a request, such as "does Wire City have 20 AWG wire in red?" using a messaging application. The system may receive the text message, parse the request to determine that a trigger event has occurred, and initiate a call to carry out an appropriate action. For example, the system may place a call to the nearest Wire City to inquire as to whether they have 20 gauge red wire in stock at the moment.

**[0102]** Similarly, in certain examples, the system **100** includes or forms a part of, or is configured to communicate with a virtual assistant system that itself is a collection of software agents for assisting a user with various services or tasks. For example, the user may input to a virtual assistant (by voice or text entry) "is my dry cleaning ready?" The virtual assistant may process this input and determine that communication with a business is required to satisfy the query, and accordingly communicate with the system to identify intent, place the call, and execute the appropriate workflow.

**[0103]** In certain examples, the system **100** autonomously executes a task through multiple dialogs with multiple humans, respectively, and may collect, analyze, take action on, and/or present individual or cumulative results of the dialogs. For example, if a task is assigned to the system **100** to gather data on when the busiest time is for a number of restaurants in a specified area, the system **100** may automatically place calls to each restaurant asking how many customers are seated over a period of time to analyze the data and provide a result.

**[0104]** FIG. 2A shows an example process **200** for completing a task assigned by a user. Briefly, the process **200** may include mapping a conversation to an initial node of a set of predefined workflows, each linked by intents (**202**), selecting an outgoing message based on a current node of the workflow (**204**), receiving a response from a human user (**206**), mapping the response to an intent within the predefined workflow (**208**), selecting a next node as the current node within the workflow based on the intent (**210**), and repeating **204-210** until an end node of the set of linked nodes in the predefined workflow is reached. The process **200** can be executed by a call inhibiting system, such as the system **100**.

**[0105]** The process **200** may include mapping a conversation to an initial node of a set of predefined workflows, each linked by intents (**202**). For example, the flow manager **132** as described above with respect to FIG. 1 can map a conversation to an initial node of a set of predefined workflows that are each linked by intents. In some examples, a conversation between the system **100** and a human callee

may be initiated by a user. In some examples, the conversation includes an intent that maps to a node of a set of predefined workflows. For example, the system **100** may store a set of predefined workflows with actions to be performed. In some examples, the system may select a predefined workflow based on the identified intent. Each of the workflows may be linked by intents. In some examples, the system **100** may place a telephone call to a business specified by the user in the conversation. In some examples, the business may be a restaurant, a salon, a doctor's office, etc. In some examples, the system may consider the call successfully placed only if a human answers, and if no one answers, or if the system is directed to a phone tree and does not successfully navigate the phone tree, the system may determine that the call was not successfully placed.

**[0106]** The process **200** may include selecting an outgoing message based on a current node of the workflow (**204**). For example, the flow manager **132** may select a message saying "Hello, I would like to schedule an appointment for a haircut," if the current node of the workflow indicates that the user would like to schedule such an appointment.

**[0107]** The process **200** may include receiving a response from a human user (**206**). For example, the system **100** may receive a response from a human callee on the other end of a telephone call, such as "Sure, and what time and date would you like to schedule this appointment for?". In some examples, the system **100** may record the response (e.g., using the session recorder **114**). In some examples, the system **100** may reproduce the response for a human operator. In some examples, a human operator may be monitoring the call (e.g., using the operator controller **134**).

**[0108]** The process **200** may include mapping the response to an intent within the predefined workflow (**208**). The flow manager **132** can map the response to an intent within the predefined workflow. In some examples, the system compares the identified intent with the intents through which the set of predefined workflows are each linked.

**[0109]** The process **200** may include selecting a next node as the current node within the workflow based on the intent (**210**). For example, the flow manager **132** may determine, using the intent, a next node of the workflow. The flow manager **132** may then designate the next node as the current node. The process **200** may include repeating **204-210** until an end node is reached. Thus, the designated current node is used in each repeated cycle of **204-210** to determine the next outgoing message until an end node is reached.

**[0110]** FIG. 2B shows an example process **250** for completing a task assigned by a user. Briefly, the process **250** may include receiving a task associated with an intent from a user (**252**), identifying the intent (**254**), selecting a predefined workflow based on the intent from among a set of predefined workflows linked by intents (**256**), following the predefined workflow (**258**), and completing the task (**260**). The process **250** can be executed by a call initiating system, such as the system **100**.

**[0111]** The process **250** may include receiving a task associated with an intent from a user (**252**). For example, a user may submit a search query "book an appointment for a haircut" to the system **100** through a user interface. In some examples, the search query may be received by the trigger module **110**, which detects that the query is a trigger event that indicates that a call should be placed to a particular callee. The task may be to book an appointment, and the intent may be to get a haircut. In some examples, the task or

the intent may not be explicitly entered. In some examples, a user may submit a task and an intent without entering a search query. The task associated with an intent may be received by a system for assisting with tasks.

[0112] The process 250 may include identifying the intent (254). For example, the system 100 may process the received task associated with an intent and identify the intent. In some examples, the intent may be explicitly input and separate from the task. In some examples, the intent may be a property of the task. In some examples, the input is provided as speech input, and the speech endpoint detector 120 provides a parsed output to the speech-to-text module 126 which sends the text to the text-to-intent module 130, which identifies an intent.

[0113] The process 250 may include selecting a predefined workflow based on the intent from among a set of predefined workflows linked by intents (256). For example, the system 100 may store a set of predefined workflows with actions to be performed. In some examples, the system may select a predefined workflow based on the identified intent from (254). For example, the flow manager 132 can select a predefined workflow based on the identified intent from (254) by the text-to-intent module 130. In some examples, the system compares the identified intent with the intents through which the set of predefined workflows are each linked.

[0114] The process 250 may include following the predefined workflow (258). For example, the system 100 may include modules that follow instructions included in the predefined workflow. In some examples, a bot of the system 100 may follow the instructions included in the predefined workflow. For example, instructions may include instructing the trigger module 110 to provide control data to the dialer 106 to place a call to and converse with a human representative of a business.

[0115] The process 250 may include completing the task (260). For example, the system 100 may complete the entire task assigned, such as paying a bill, changing a reservation for dinner, etc. In some examples, the system 100 may complete a portion of the task, such as placing a call and navigating a phone tree until it reaches a human. In some examples, the system 100 may complete a portion of the task specified by the user. For example, the user may specify that the system complete all of the task and transfer the call to the user for verification.

[0116] Many use cases may involve users who want to purchase something from a business, but have trouble doing so due to complexity, menu navigation, language challenges, reference knowledge, etc., required for the transaction. Transaction queries may garner support from humans on the vendor side willing to help the system succeed in order to complete the transaction. In some examples, the system provides critical assistance in developing countries and low tech and service industries, such as plumbing, roofing, etc. Workflows may be employed to both assist a human user to successfully navigate such transactions as well as encourage the vendor-side system to assist the user. The system is scalable to accommodate varying use cases. For example, a restaurant booking application may partner with thousands of businesses worldwide; the system disclosed herein can be configured to issue a restaurant booking at whatever scale is required.

[0117] FIG. 3 shows an example workflow 300 for a process executed by the system. In this particular example,

a simple Boolean question is asked by the bot of the system 100. It is understood that the system is able to respond to questions of higher complexity, and that the work flow 300 is presented for simplicity of explanation.

[0118] The flow 300 shows an example question posed by the bot: "Are you open tomorrow?" The possible responses provided by the human are laid out, and the bot's responses to each of the human's responses are provided. Depending on the human's response, there are several stages of the flow 300 to which the system 100 could be directed. Stages shown in double-borders are end stages in which system 100 exits the flow 300. For example, in response to the binary question posed by the bot, the human callee could confirm that the business is open tomorrow, ending the flow 300. The human callee could confirm that the business is not open tomorrow, ending the flow 300. The human callee could ask the bot to hold, thus sending the bot into a hold flow separate from the flow 300 and ending the flow 300.

[0119] For ease of access to users and to promote propagation of the system 100, the system 100 is integrated with existing applications, programs, and services. For example, the system 100 may be integrated with existing search engines or applications on a user's mobile device. Integration with other services, or verticals, allow users to easily submit requests for tasks to be completed. For example, the system 100 may be integrated with a search engine knowledge graph.

[0120] In some use cases, real-time judgment of a human might be automated. For example, the system 100 may automatically detect that the user is running ten minutes late to a barber shop appointment and alert the barber prior to the user's arrival.

[0121] The system 100 can select certain parameters of the bot based on the context of the conversation being conducted or data about a particular callee stored within a knowledge database. For example, the system 100 can determine based on a callee's accent, location, and other contextual data, that the callee is more comfortable with a language different than the language the call is currently being conducted in. The system 100 can then switch to the language with which the bot believes the callee is more comfortable and ask the callee if he or she would prefer to conduct the call in the new language. By mirroring certain speech characteristics of the human callee, the system 100 increases the possibility of a successful call. The system 100 reduces potential sources of friction within a conversation due to speech characteristics to reduce the strain accumulated during a call. These characteristics can include average length of word used, complexity of sentence structure, length of pauses between phrases, the language the callee is most comfortable speaking, and various other speech characteristics.

[0122] FIG. 4 is a block diagram 400 of the call triggering module of the system 100. The trigger module 110 is communicably connected to the dialer 106, and provides instructions to the dialer 106 to initiate calls to particular callees or sets of callees based on detecting a trigger event. In some examples, the trigger module 110 can communicate with the flow manager 132 to provide trigger event data that the flow manager 132 uses to select a node of a particular workflow or provide instructions to the dialer 106.

[0123] The trigger module 110 receives input from various modules, including an inconsistency detector 402, a third party API 404, a trend detector 406, and an event identifier

**408.** The trigger module **110** can also receive input from the flow manager **132**. In some examples, each of the modules **402-408** are integral to the system **100**. In other examples, one or more of the modules **402-408** are remote from the system **100**, and are connected to the trigger module **110** over a network, such as a local area network (LAN), a wide area network (WAN), the Internet, or a combination thereof. The network can connect one or more of the modules **402-408** to the trigger module, and can facilitate communication between components of the system **100** (e.g., between the speech API **128** and the speech-to-text module **126**).

**[0124]** The inconsistency detector **402** receives data from multiple different sources and detects inconsistencies between data values from a first data source and corresponding data values from a second source. For example, the inconsistency detector **402** can receive data indicating the operating hours of a clinic and detect that the operating hours for a clinic listed on the clinic's website are different from the operating hours posted outside of the clinic. The inconsistency detector **402** can provide data to the trigger module **110** indicating the source of contention, the type of the data value in which the inconsistency was detected, the data values that are in conflict, and various other characteristics. In some examples, the inconsistency detector **402** provides the trigger module **110** with instructions to initiate a call to a particular callee. In other examples, the trigger module **110** determines, based on the data received from the inconsistency detector **402**, a particular callee to contact and fields of data to be collected from the particular callee.

**[0125]** The trigger module **110** can detect a trigger event based on data provided by the inconsistency detector **402**. Trigger events can include receiving user input indicating a discrepancy. For example, the trigger module **110** can receive user input through a user interface **410**. The user interface **410** can be an interface for a separate application or program. For example, the user interface **410** can be a graphical user interface for a search engine application or a navigational application.

**[0126]** In some implementations, the user interface **410** can prompt a user to provide information. For example, a user is detected to be at a store after the store's advertised closing hours, the system can ask the user if the store is still open or ask the user to input the hours. The user can input the requested data through the user interface **410**, and the inconsistency detector **402** can determine whether there exists a discrepancy between the data input through the user interface **410** and corresponding data from a second source, such as a knowledge base **412**. The knowledge base **412** can be a storage medium such as a remote storage device, a local server, or any of various other types of storage media. The inconsistency detector **402** can determine whether the user is at the store a predetermined amount of time outside of regular hours (e.g., more than 20 minutes, because stores may stay open for a few extra minutes for a particularly slow customer).

**[0127]** In another exemplary situation, the inconsistency detector **402** can determine that information on an organization's website is outdated. For example, the inconsistency detector **402** can detect, based on data from the knowledge database **412**, that a bass fishing club's website shows that there are monthly meetings on the first Wednesday of every month, but that the club's more active social media profiles all indicate that monthly meetings occur on the second

Tuesday of every month. The inconsistency detector **402** can then output to the trigger module **110** data indicating this detected inconsistency.

**[0128]** Trigger events can include determining that a particular set of data has not been updated for a predetermined amount of time. For example, a user of the system **100** can specify an amount of time after which data should be refreshed regardless of any other trigger events occurring. The inconsistency detector can compare a last-updated timestamp for a particular data value and determine, based on the timestamp, whether the predetermine amount of time has passed. Characteristics for particular data fields, including timestamps and the data values themselves, can be stored in the knowledge database **412**. A timer **414** can provide data to the knowledge database **412** to update the amount of time that has passed. The inconsistency detector **402** can determine that a predetermined period of time has passed based on the timing data provided by the timer **414**.

**[0129]** For example, the inconsistency detector **402** can determine, based on data from the knowledge database **412**, that operating hours for small coffee shops in Ithaca, N.Y. have not been updated for three months. The inconsistency detector **402** can then provide output data to the trigger module **110** indicating the detected event.

**[0130]** Trigger events can include receiving requests to initiate a call from one or more users. For example, the trigger module **110** can detect receiving a request from a user through the third party API **404**. The third party API **404** is communicably connected to a user interface, such as the user interface **416**, through which a user can provide input indicating a request to initiate a call. For example, the user interface **416** can be a graphical user interface for an application through which a user can request call campaigns to be scheduled and executed. A user can provide data indicating a particular callee or set of callees, and the particular data requested for extraction. For example, a user can request that a call campaign to each hardware store in Virginia that sells livestock supplies be conducted and that the hardware stores are asked whether they provide chick starter feed (e.g., such that an index of locations carrying the supplies is available for later search).

**[0131]** During the call, the callee can schedule a different time for the system **100** to call the callee back. For example, if asked about whether any changes have been made to a menu of a restaurant, a human callee can ask the system **100** to call them back in an hour or the next day, after they've had a chance to look over the new menu, for further action. The system **100** can then schedule a call for the requested time. In some examples, the trigger module **110** can schedule a trigger event for the future. In other examples, the flow manager **132** can schedule an intent or a call event to be executed by the dialer **106** to initiate the call.

**[0132]** Trigger events can include trends or patterns detected in stored data within a knowledge database or data provided in real time. For example, trends detected in search data received from a search engine **418** can be trigger events. Search engine **418** receives search requests from users, and can provide data indicating the search requests to the trend detector **406**. The trend detector **406** analyzes the data received and detects trends in the received data. For example, if searches for Cuban restaurants in Asheville, N.C. have increased 500% in the past month, the trend detector **406** can detect the increase in searches and provide data indicating the trend to the trigger module **110**.

[0133] The trend detector 406 can output data to the trigger module 110 that indicates a particular callee or a set of callees based on the identified trend. In some implementations, the trend detector 406 provides data indicating the detected trend and the trigger module 110 determines a particular callee or set of callees based on the identified trend. For example, the trend detector 406 can determine that there has been a 40% increase in searches for “tornado Lincoln Nebr.” and provide the keywords of the search to the trigger module 110. The trigger module 110 can then determine that calls should be placed to all stores that provide emergency supplies to ascertain how much stock each store has of essential items, and their hours of operation (e.g., for indexing and later search by users of the search engine).

[0134] Trigger events can include particular events of interest that have been identified as having effects on normal operations of businesses, organizations, individual persons, etc. The event identifier 408 receives data from various third party sources, including a third party database 420 and an event database 422. The event identifier 408 can receive data from other sources, such as local memory devices or real time data streams. The event identifier 408 identifies particular events from the databases 420 and 422 and outputs data indicating the identified events to the trigger module 110. In some examples, the trigger module 110 selects a particular callee or set of callees and data to be requested during a call based on the data provided by the event identifier 408.

[0135] Particular events that can affect operation of businesses, organizations, and individual persons include extreme weather conditions, federal holidays, religious holidays, sporting events, and various other happenings.

[0136] The third party database 420 provides the event identifier 408 with data from various third party data sources, including the weather service, governmental alerts, etc. For example, the third party database 420 can provide the trend identifier 408 with storm warnings. The event identifier 408 can then determine that a winter storm is approaching the north-eastern corner of Minneapolis, Minn., and can determine that calls should be placed to hardware stores within the north-eastern corner of Minneapolis to determine the current stock of generators available.

[0137] The event database 422 provides the event identifier 408 with data from various data sources and specifically includes data indicating known events. For example, the event database 422 can provide data indicating federal and state holidays, religious holidays, parades, sporting events, exhibition openings, visiting dignitaries, and various other events.

[0138] For example, if a particular city is hosting the Super Bowl, the event database 422 can provide the data to the event identifier 408, which provides data indicating the event to the trigger module 110. The trigger module 110 can determine, based on known information about the current Super Bowl and stored information about the past Super Bowl, that calls should be placed to all hotels in the area to confirm availability and pricing. The trigger module 110 can also determine that calls should be placed to sporting goods stores to determine the availability of jerseys for each of the teams participating in the Super Bowl. In such situations, other information that affects operation of businesses, organizations, or individual persons that the trigger module 110 can request includes closings of office buildings and schools,

changes to public transportation schedules, special restaurant offerings, or various other information.

[0139] One or more of the various modules of the system 100 can determine, based on the received event information from the event identifier 408, inferred trigger events or information to request. For example, if it is Dia de Muertos, South American restaurants, and Mexican restaurants in particular, may have special menus or hours because of the celebrations. In such examples, the trigger module 110 can provide instructions to the dialer 106 to conduct calls to South American restaurants to update operating hours for the day and the menu.

[0140] In some implementations, a trigger event can be detected from the call placed by the system 100 itself. The flow manager 132 can determine, based on portions of the conversation conducted by the system 100, that an intent has been expressed during the conversation suggesting that a call should be placed. For example, if a human callee says “Yes, we are still open until 8 p.m. every Thursday, but next week we will be switching over to our summer schedule, and we will be open until 9:30 p.m.,” then the flow manager 132 can identify an intent that provides further information regarding a data field.

[0141] In some implementations, a trigger event can include receiving unsatisfactory results from a previously placed call. For example, if the bot places a call to a business to determine if the business will have special holiday hours over the Independence Day holiday, and does not have at least a threshold amount of confidence in the veracity of the answer provided by the human representative of the business, the system 100 can schedule a call for another specific day or time, such as July 1, to determine whether special holiday hours will be in place. In such examples, the trigger module 110 can schedule a trigger event or provide information to the flow manager 132 to schedule an action. In some examples, the flow manager 132 schedules initiating a callback by scheduling transmission of instructions to the dialer 106.

[0142] The system 100 has a common sense module 133 that allows the flow manager 132 to schedule and select nodes of a particular workflow intelligently. For example, in the above situation, when there is a deadline for the usefulness of the information being requested during the call, the common sense module 133 can also determine when to schedule the call and what information to request. In some examples, the common sense module 133 is a component of the flow manager 132, as described in FIG. 1. In other examples, the common sense module 133 is a component of the trigger module 110 and facilitates the trigger module 110 making intelligent determinations as to whether a call should be initiated.

[0143] FIG. 5 shows an example process 500 for initiating a phone call. Briefly, the process 500 may include receiving, by a call triggering module of a call initiating system for placing calls and conducting conversations between a bot of the call initiating system and a human callee during the calls, data indicating a first event (502), determining, by the call triggering module and using the data indicating the first event, that the first event is a trigger event that triggers a workflow for the call initiating system that begins with initiating a phone call (504), selecting, based on the determined trigger event, a particular workflow (506), and in response to the selecting, initiating a phone call to a callee specified by the particular workflow (508).

[0144] The process 500 may include receiving, by a call triggering module of a call initiating system for placing calls and conducting conversations between a bot of the call initiating system and a human callee during the calls, data indicating a first event (502). For example, the trigger module 110 can receive data from the inconsistency detector 402 indicating a discrepancy between the business hours of Sally's Saloon of Sweets posted on the store's website and the business hours stored in a search index relating to that business.

[0145] The process 500 may include determining, by the call triggering module and using the data indicating the first event, that the first event is a trigger event that triggers a workflow for the call initiating system that begins with initiating a phone call (504). In some examples, the determined trigger event is an inconsistency of a value associated with a first data source and a corresponding value associated with a second data source. For example, the trigger module 110 can use the detected inconsistency from the inconsistency detector 402 to determine that the inconsistency is a trigger event that will trigger a workflow to determine what the actual business hours are for Sally's Saloon.

[0146] In some examples, the data indicating the first event is provided by a user. For example, a user can report that there is a discrepancy between the hours posted on the website for Sally's Saloon and the hours posted on the storefront for Sally's Saloon.

[0147] In some examples, the determined trigger event is a user request. For example, a user can provide input through a user interface, such as the user interface 416, to a third party API, such as the third party API 404, to request scheduling and execution of calls to a particular callee or a set of particular callees.

[0148] In some examples, the determined trigger event is a particular type of event that is one of: a weather event, a sporting event, an entertainment event, or a seasonal event. For example, the event identifier 408 can determine that the Head of the Charles regatta is occurring in Boston, Mass., and can provide the event data to the trigger module 110. The trigger module 110 can then determine that the regatta is a trigger event.

[0149] In some examples, the determined trigger event is a trend detected in search requests submitted to a search engine. For example, the trend detector 406 can receive search engine data from the search engine 418 and determine that Spanish tapas restaurants are trending. The trend detector 406 can provide data indicating the trend to the trigger module 110 and the trigger module 110 can determine that the trend is a trigger event.

[0150] In some examples, the determined trigger event is the passing of a predetermined period of time. For example, the inconsistency detector 402 can determine, based on data in the knowledge database 412 from the timer 414, that the menu for the Cuban restaurants in Manhattan, N.Y. have not been updated for four months. The inconsistency detector 402 can provide the timing data to the trigger module 110, and the trigger module 110 can determine that the passing of four months without updating the menu data for the Cuban restaurants in Manhattan is a trigger event. The trigger module 110 can then provide data to the flow manager 132 suggesting that the Cuban restaurants in Manhattan be called to obtain updated menu information.

[0151] The process 500 may include selecting, based on the determined trigger event, a particular workflow (506).

The trigger module 110 can provide the trigger event data to the dialer 106 or to the flow manager 132 to use in selecting a particular workflow or node of a workflow. For example, the trigger module 110 can provide the trigger event data indicating the inconsistency in posted business hours for Sally's Saloon for Sweets to the flow manager 132, which uses the data to select a particular workflow to call Sally's Saloon to resolve the discrepancies.

[0152] The process 500 may include initiating, in response to the selecting, a phone call to a callee specified by the particular workflow (508). The flow manager 132 can provide instructions to the dialer 106 indicating a particular callee to be contacted. For example, the flow manager 132 can provide instructions to the dialer 106 to call Sally's Saloon.

[0153] While the initiation of workflows, and more specifically the placement of calls, by the systems and methods described herein may be relatively automated by triggering events, safeguards may be included in system 100 to prevent unwanted calls or calls in violation of local regulations. For example, if a callee indicates that they no longer wish to receive calls from the system, the system may note this and build in checks for calls to the callee's number to prevent further calls.

[0154] Furthermore, to the extent that the systems and methods described herein collect data, the data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed or permanently obscured. For example, identity of the callee may be permanently removed or treated so that no personally identifiable information can be determined, or a callee's geographical location may be generalized where location information is obtained, so that a particular location of a user cannot be determined where appropriate. Where personal, private, or sensitive information is received during a call, whether requested as part of the workflow, volunteered by the callee, or inadvertently received, workflows may include steps for permanent removal or obfuscation of that information from the system.

[0155] In certain examples, the system 100 may provide the user with the current status of efforts to execute the request for assistance a task, either automatically or upon user request. For example, the system 100 may provide the user with a status of a task being carried out through a notification on the device the user is using, such as a computer, a mobile device, etc. In some examples, the system 100 may notify the user of the status of an ongoing task through other means, such as a messaging application, through telephonic communication, etc.

[0156] FIG. 6 is a block diagram 600 of the task manager module of the system 100. The task manager module 140 is connected to the communication platform 102, the trigger module 110, the task information storage 150, and the sessions storage 116. When a user communicates a task through the communication platform, the task information is stored in the task information storage 150 and the task manager module 140 determines when the task should be scheduled. The task manager may associate the task with a trigger event. The task may have a status that is initially set to "new" or some other indicator that there has been no processing on the request. Once the trigger event occurs, the trigger module 110 will start the dialing process. In some implementations, the task manager module 140 monitors the

sessions storage to update the status of each task when the status of the tasks change from initiated, to in progress, to complete.

**[0157]** From the sessions information, the task manager module can determine the status and outcome of each call. For example, a bot may try to call a restaurant several times before getting through to someone to make a reservation. The sessions storage holds information about each call that the bot makes. In some implementations, the task manager module may periodically poll the session storage to determine the status of a call task, i.e., whether a call has been initialized, is in progress, or has been complete. In other implementations, the sessions storage may send the outcome of a call to the task manager module in order to update the status of the task in the task information storage.

**[0158]** In some implementations, a call is reviewed by an operator through an operator dashboard that displays call tasks and information about the progress of the tasks.

**[0159]** FIG. 7A illustrates an operator dashboard that shows information about progress of existing call tasks. For example, FIG. 7A shows tasks for haircut appointments. The operator dashboard may provide information about the appointments including: the booking time, requester's name, service requested, business name, date, and time of the appointment. An operator may be able to review the requests and the associated sessions information from calls associated with the requests to determine if the requested appointments have been properly booked.

**[0160]** FIG. 7B illustrates an operator review screen for reviewing one of the user requested tasks. The screen may show the operator the current status of the task. As illustrated in FIG. 7B, the task is complete since the reservation has been made. However, in some cases, the task may not be complete and the reservation may not have been made. The operator may have the option of playing the recording associated with the task or looking at other stored information from the call, e.g., transcriptions, extracted intents, etc., calling the business associated with the task, or scheduling an automated call for the future. Additionally, the operator may have the option of providing the current status of the task to the requesting user.

**[0161]** A user can also request the status of a task through the communication platform **102**. Additionally or alternatively, the task manager module **140** can determine when to send the user a status update based on other triggering events such as task status changes or time.

**[0162]** FIG. 8 is a flow diagram that shows an example of a process **800** for providing the status of a task. The process **80** may include determining, by a task manager module, that a triggering event has occurred to provide a current status of a user call request (**802**). As discussed above, triggering events may include: a user request for status, a passage of a certain amount of time, or a change in the status of a particular task. The process **800** then includes determining, by the task manager module, the current status of the user call request (**804**). The task manager module can determine the current status by checking the status in the task information storage. The status of a task is initialized when the task is added to the task information storage **150**. As calls associated with the task are made and completed, the status of the task is updated. The task manager then generates a representation of the current status of the user call request (**806**). The representation can be a visual or an audio representation that conveys the current status of the task.

The process **800** provides the user with the generated representation of the current status of the user call request (**808**).

**[0163]** FIG. 9A shows a visual status of the haircut appointment request of FIG. 1B while the appointment scheduling is in progress. A user may be able to access a user interface to check the status of a task request or the status may be sent to a user device such as a smart phone, smart watch, laptop, personal home assistant device, or other electronic device. The status may be sent by email, SMS, or other mechanism.

**[0164]** FIG. 9B shows a visual status of the haircut appointment request of FIG. 1B once the appointment has been successfully scheduled. This status may be requested by the user or may be sent to the user without the user prompting once the appointment has been successfully booked.

**[0165]** FIG. 10A illustrates an oral status request and update of the restaurant reservation request of FIG. 1C. As illustrated in FIG. 10A, in response to the user asking whether the restaurant reservation has been made, the system may explain the steps it has taken to complete the task, such as calling the restaurant twice. The system may also tell the user when the next time the system is scheduled to attempt the call and may notify the user of the status after the call attempt.

**[0166]** FIG. 10B illustrates an oral status update provided by the system without prompting by the user for the restaurant reservation request of FIG. 1C. Once the system knows that the user's task has been complete, the system can provide the user with a status update. In some implementations, the system provides the user with the status update immediately. In other implementations, the system determines a convenient time or method for notifying the user. For example, the user may request a dinner reservation in London, United Kingdom. However, the user may presently be located in Mountain View, Calif., USA. The system may attempt to call the restaurant at a time when the user is sleeping. If the system confirms the reservation at 12 pm in London, the system may determine that sending a status update text message at 4 am PDT may wake the user. The system may then choose an alternate status update method, i.e., email, or hold the status update for a more convenient time for the user. The system can determine the appropriate, convenient times and method to provide status updates to the user using information from the user's schedule, time zone, habits, or other personal information of the user.

**[0167]** In some implementations, the system may use user information to determine the urgency of a task or whether to repeat an effort to complete a task. For example, the system may be trying to make a reservation for a user at a specific restaurant in Mountain View, Calif. The user's travel to Mountain View may end on May 15th. If the system is still not successful on May 15th, it does not make sense for the system to continue to request reservations for May 16th or later since the user's trip will be over. However, it does make sense to call twice as often on May 14th as compared with earlier dates in order to get a hold of someone at the restaurant to make a reservation. A task may become more urgent as a deadline approaches and less urgent or obsolete as the deadline passes.

**[0168]** In some implementations, the bailout module **136** of FIG. 1B determines the type of intervention that should be introduced for a call while the call is in progress. The

bailout module **136** may choose to manually bailout of a bot conversation in real-time and explain that another will be taking over the call. In other implementations, the module may allow a human operator to silently take over the call. Additionally or alternatively, the bailout module **136** may choose to politely end a phone call between a bot and a human with no manual intervention.

**[0169]** FIG. 11 shows an example process **1100** for transitioning a phone call from a bot to a human. The process **1100** may include, analyzing, by a call initiating system, a real-time conversation between a first human and the bot during a phone call between the first human on the first end of the phone call and the bot on a second end of the phone call (**1102**). The call initiating system may then determine based on analysis of the real-time conversation, whether the phone call should be transitioned from the bot to a second human on the second end of the phone call (**1104**). In response to determining that the phone call should be transitioned to a second human on a second end of the phone call, transitioning, by the call initiating system, the phone call from the bot to the second human (**1106**).

**[0170]** In order to determine the type of intervention that is most appropriate for a particular bot phone call, the bailout module **136** may identify strain events or look for other indications that the call should either be terminated or handed over to a human operator.

**[0171]** In some implementations, the bailout module **136** identifies strain events that indicate strain on the human or on the bot to appropriately respond to the human's questions. Each time the bailout module **136** identifies a strain event, it increases the stored levels of both local and global strain of the call. Whenever the conversation seems to get back on track, the bailout module **136** resets the local strain level. For example, the human may ask the bot, "How many highchairs will your party need?" when the bot has called a restaurant to make reservations for a party of 6. The bot may respond with, "we all need chairs." The human may have a slightly irritated tone based on the bot's response and respond, "Yes, I know you all need chairs, but how many high chairs for babies do you need?" The system may detect intonation patterns, i.e., a higher pitch at the beginning of a human's statement, at the end of the statement, or throughout the statement. In some implementations, intonation patterns are pre-associated with stress or irritation. The system can match the pre-associated patterns with the pattern detected in the real-time conversation. In some implementations, intonation patterns can detect repeated words, talking intentionally slower, or keywords or phrases ("are you even listening to me?" "am I talking to a robot?")

**[0172]** When the system detects a slightly irritated tone of the human it increases the local strain level of the call. The local strain is a running score that reflects the likely amount of strain associated with the current state. If any of the strain indicators appear in a human statement within the real-time conversation, the strain score rises until the score reaches an intervention threshold. If none of the stress indicators appear, the system may indicate that the call is proceeding according to the workflow and the local strain score decreases or remains low (or 0). If the bot appropriately responds to the question by providing a response expected by the human, such as, "We have no children in our party," the system can decrease the local strain. If the system detects that the human responds with no irritation in his or her voice,

the bailout module may determine that the call is back on track and reset the local strain to a default value or reset to zero.

**[0173]** The global strain for a phone call only accumulates up. Whereas the local strain tries to evaluate whether the current correspondence with the human is strainful, the global strain tries to assess the total strain for the entire call. For example, a threshold may be set for three misunderstandings before the bot bails out to a human operator. If the bot didn't understand the human three times in a row, the local strain would be high and would cause the bot to bailout. In a different call, if the bot didn't understand the other side twice in a row, but did understand the third sentence, the local strain would get reset on the third interaction, and the conversation would probably continue. The global strain would still maintain information to indicate that there were two misunderstandings between the bot and the human. If at a later time during the call, the bot doesn't understand the human twice in a row again, the global strain level would go above a threshold and the bot will probably bail out, even though the local strain is still below the set threshold of three misunderstandings.

**[0174]** As discussed above, if either the local or global strain levels reach some threshold, the bailout module **136** will indicate to the system **100** that it is time for manual intervention or to politely exit from the call. In some examples, the bailout module **136** would consider an event as a strain event whenever it needs to repeat itself, apologize, ask for clarification, etc., as well as when the human corrects the system **100** or complains about the call (e.g., "I can't hear you, can you hear me?").

**[0175]** In some examples, the bailout module **136** would consider an event a strain event if the human asks if the bot is a robot, mocks the bot, i.e., by asking nonsensical questions, or acts in some other way that the system is not expecting, (e.g., if the system gets asked about a sporting event when trying to make a reservation for a restaurant.)

**[0176]** In some implementations, the bailout module **136** is a feature-based rule set that determines when the system should bailout to manual intervention. One feature-based rule may be a rule that states that when two consecutive unknown input intents occur, the system should bailout. A different rule could state that when four unknown input intents occur anywhere during the call, the system bails out to a manual operator. The system keeps track of events occur in the conversation and determines whether events have happened that meet the criteria of a rule.

**[0177]** In other implementations, the bailout module **136** uses machine learning to predict when to bailout to a human operator automatically. For example, the bailout module **136** can receive, as input to one or more machine learning models, intents from a conversation with a human. The machine learning models can determine, based on the received intents and historical intents and outcomes, whether to bail out to a human operator. The system can train the machine learning model on features from annotate recordings that denote when a bailout should have occurred or should not have occurred. The machine learning module can then predict when a bailout should likely occur given a set of input features.

**[0178]** The bailout module **136** uses many factors to determine bailout including: the conduct of the human, the

tone of the human, the determined annoyance level of the human, the language the human uses, or the word choice of the human.

**[0179]** The system **100** can escalate a conversation being conducted by the bot to a human operator to handle. For example, if there is a threshold amount of strain in a particular conversation, the bailout module **136** can provide feedback data to the flow manager **132**. The flow manager **132** may instruct the bot to hand the call over to a human operator that provides input through the operator controller **134**, with or without audibly alerting the human callee. For example, the bot can say “Sure, thank you for your time today. Here is my supervisor.” A human operator can then complete the task that the bot was attempting to carry out through the operator controller **134**.

**[0180]** The bailout module **136** can also determine a confidence level, which defines the confidence that the system has in the current task being accomplished. For example, the bot may be tasked with making dinner reservations for a user. If the bot calls the restaurant and the human asks multiple questions for which the bot does not know the answer, the system may have a low confidence in the current task being accomplished. After the system receives a question for which the system does not have an answer, the system’s confidence level in accomplishing the task may go lower. If the system recovers and the system determines that the conversation is moving towards accomplishing a task, the system may raise the confidence level.

**[0181]** In some implementations, the system hands off the phone conversation to a human operator who monitors calls. The system may alert an operator of the need to transition the phone call using an operator user interface or some other notification mechanism. Once notified, the operator may have a finite time to transition the phone call before the system determines to terminate the call. The system may use the same voice as the operator. In such cases, the transition from a bot to the operator can be transparent for the other side as the voice remains the same.

**[0182]** In other implementations, the system hands off the phone conversation to the human user who requested the task. The system can alert the user of the in-progress phone call. The system can let the user know when there is a problem with completing the task or when the bot has been asked a question to which the bot does not know the answer. The bot may text, email, or in some other way communicate the details of the conversation for which the bot needs user input. In some implementations, the bot will wait a threshold amount of time, i.e., 5 seconds, for the user to respond before continuing the conversation without user input. Since the conversation is happening in real-time, the bot cannot wait a long period of time for user response. In some implementations, the system may try to transition the phone call to the requesting user when the system has determined that the phone call needs to be transitioned away from the bot. As discussed above, the system may wait a threshold amount of time for the user to respond and take over the phone call. In some implementations, if the user does not take over the phone call in the threshold amount of time, the system will transition the phone call to an operator. In other instances, the system will end the phone conversation. The system may also use the same voice as the human user so that transition from the bot to the user is seamless from the other side of the conversation.

**[0183]** FIG. 12 shows an example of a computing device **1200** and an example of a mobile computing device **1250** that can be used to implement the techniques described above. The computing device **1200** is intended to represent various forms of digital computers, such as laptops, desktops, workstations, personal digital assistants, servers, blade servers, mainframes, and other appropriate computers. The mobile computing device **1250** is intended to represent various forms of mobile devices, such as personal digital assistants, cellular telephones, smart-phones, and other similar computing devices. The components shown here, their connections and relationships, and their functions, are meant to be exemplary only, and are not meant to limit implementations of the inventions described and/or claimed in this document.

**[0184]** The computing device **1200** includes a processor **1202**, a memory **1204**, a storage device **1206**, a high-speed interface **1208** connecting to the memory **1204** and multiple high-speed expansion ports **1210**, and a low-speed interface **1212** connecting to a low-speed expansion port **1214** and the storage device **1206**. Each of the processor **1202**, the memory **1204**, the storage device **1206**, the high-speed interface **1208**, the high-speed expansion ports **1210**, and the low-speed interface **1212**, are interconnected using various busses, and may be mounted on a common motherboard or in other manners as appropriate.

**[0185]** The processor **1202** can process instructions for execution within the computing device **1200**, including instructions stored in the memory **1204** or on the storage device **1206** to display graphical information for a GUI on an external input/output device, such as a display **1216** coupled to the high-speed interface **1208**. In other implementations, multiple processors and/or multiple buses may be used, as appropriate, along with multiple memories and types of memory. Also, multiple computing devices may be connected, with each device providing portions of the necessary operations (e.g., as a server bank, a group of blade servers, or a multi-processor system).

**[0186]** The memory **1204** stores information within the computing device **1200**. In some implementations, the memory **1204** is a volatile memory unit or units. In some implementations, the memory **1204** is a non-volatile memory unit or units. The memory **1204** may also be another form of computer-readable medium, such as a magnetic or optical disk.

**[0187]** The storage device **1206** is capable of providing mass storage for the computing device **1200**. In some implementations, the storage device **1206** may be or contain a computer-readable medium, such as a floppy disk device, a hard disk device, an optical disk device, or a tape device, a flash memory or other similar solid state memory device, or an array of devices, including devices in a storage area network or other configurations. A computer program product can be tangibly embodied in an information carrier. The computer program product may also contain instructions that, when executed, perform one or more methods, such as those described above. The computer program product can also be tangibly embodied in a computer- or machine-readable medium, such as the memory **1204**, the storage device **1206**, or memory on the processor **1202**.

**[0188]** The high-speed interface **1208** manages bandwidth-intensive operations for the computing device **1200**, while the low-speed interface **1212** manages lower bandwidth-intensive operations. Such allocation of functions is



exemplary only. In some implementations, the high-speed interface **1208** is coupled to the memory **1204**, the display **1216** (e.g., through a graphics processor or accelerator), and to the high-speed expansion ports **1210**, which may accept various expansion cards (not shown). In the implementation, the low-speed interface **1212** is coupled to the storage device **1206** and the low-speed expansion port **1214**. The low-speed expansion port **1214**, which may include various communication ports (e.g., USB, Bluetooth, Ethernet, wireless Ethernet) may be coupled to one or more input/output devices, such as a keyboard, a pointing device, a scanner, or a networking device such as a switch or router, e.g., through an network adapter.

[0189] The computing device **1200** may be implemented in a number of different forms, as shown in the figure. For example, it may be implemented as a standard server, or multiple times in a group of such servers. In addition, it may be implemented in a personal computer such as a laptop computer **1222**. It may also be implemented as part of a rack server system **1224**. Alternatively, components from the computing device **1200** may be combined with other components in a mobile device (not shown), such as a mobile computing device **1250**. Each of such devices may contain one or more of the computing device **1200** and the mobile computing device **1250**, and an entire system may be made up of multiple computing devices communicating with each other.

[0190] The mobile computing device **1250** includes a processor **1252**, a memory **1264**, an input/output device such as a display **1254**, a communication interface **1266**, and a transceiver **1268**, among other components. The mobile computing device **1250** may also be provided with a storage device, such as a micro-drive or other device, to provide additional storage. Each of the processor **1252**, the memory **1264**, the display **1254**, the communication interface **1266**, and the transceiver **1268**, are interconnected using various buses, and several of the components may be mounted on a common motherboard or in other manners as appropriate.

[0191] The processor **1252** can execute instructions within the mobile computing device **1250**, including instructions stored in the memory **1264**. The processor **1252** may be implemented as a chipset of chips that include separate and multiple analog and digital processors. The processor **1252** may provide, for example, for coordination of the other components of the mobile computing device **1250**, such as control of user interfaces, applications run by the mobile computing device **1250**, and wireless communication by the mobile computing device **1250**.

[0192] The processor **1252** may communicate with a user through a control interface **1258** and a display interface **1256** coupled to the display **1254**. The display **1254** may be, for example, a TFT (Thin-Film-Transistor Liquid Crystal Display) display or an OLED (Organic Light Emitting Diode) display, or other appropriate display technology. The display interface **1256** may comprise appropriate circuitry for driving the display **1254** to present graphical and other information to a user. The control interface **1258** may receive commands from a user and convert them for submission to the processor **1252**. In addition, an external interface **1262** may provide communication with the processor **1252**, so as to enable near area communication of the mobile computing device **1250** with other devices. The external interface **1262** may provide, for example, for wired communication in some

implementations, or for wireless communication in other implementations, and multiple interfaces may also be used.

[0193] The memory **1264** stores information within the mobile computing device **1250**. The memory **1264** can be implemented as one or more of a computer-readable medium or media, a volatile memory unit or units, or a non-volatile memory unit or units. An expansion memory **1274** may also be provided and connected to the mobile computing device **1250** through an expansion interface **1272**, which may include, for example, a SIMM (Single In Line Memory Module) card interface. The expansion memory **1274** may provide extra storage space for the mobile computing device **1250**, or may also store applications or other information for the mobile computing device **1250**. Specifically, the expansion memory **1274** may include instructions to carry out or supplement the processes described above, and may include secure information also. Thus, for example, the expansion memory **1274** may be provided as a security module for the mobile computing device **1250**, and may be programmed with instructions that permit secure use of the mobile computing device **1250**. In addition, secure applications may be provided via the SIMM cards, along with additional information, such as placing identifying information on the SIMM card in a non-hackable manner.

[0194] The memory may include, for example, flash memory and/or NVRAM memory (non-volatile random access memory), as discussed below. In some implementations, a computer program product is tangibly embodied in an information carrier. The computer program product contains instructions that, when executed, perform one or more methods, such as those described above. The computer program product can be a computer- or machine-readable medium, such as the memory **1264**, the expansion memory **1274**, or memory on the processor **1252**. In some implementations, the computer program product can be received in a propagated signal, for example, over the transceiver **1268** or the external interface **1262**.

[0195] The mobile computing device **1250** may communicate wirelessly through the communication interface **1266**, which may include digital signal processing circuitry where necessary. The communication interface **1266** may provide for communications under various modes or protocols, such as GSM voice calls (Global System for Mobile communications), SMS (Short Message Service), EMS (Enhanced Messaging Service), or MMS messaging (Multimedia Messaging Service), CDMA (code division multiple access), TDMA (time division multiple access), PDC (Personal Digital Cellular), WCDMA (Wideband Code Division Multiple Access), CDM252000, or GPRS (General Packet Radio Service), among others. Such communication may occur, for example, through the transceiver **1268** using a radio-frequency. In addition, short-range communication may occur, such as using a Bluetooth, Wi-Fi, or other such transceiver (not shown). In addition, a GPS (Global Positioning System) receiver module **2570** may provide additional navigation- and location-related wireless data to the mobile computing device **1250**, which may be used as appropriate by applications running on the mobile computing device **1250**.

[0196] The mobile computing device **1250** may also communicate audibly using an audio codec **1260**, which may receive spoken information from a user and convert it to usable digital information. The audio codec **1260** may likewise generate audible sound for a user, such as through

a speaker, e.g., in a handset of the mobile computing device **1250**. Such sound may include sound from voice telephone calls, may include recorded sound (e.g., voice messages, music files, etc.) and may also include sound generated by applications operating on the mobile computing device **1250**.

**[0197]** The mobile computing device **1250** may be implemented in a number of different forms, as shown in the figure. For example, it may be implemented as a cellular telephone **1280**. It may also be implemented as part of a smart-phone **1282**, personal digital assistant, tablet computer, wearable computer, or other similar mobile device.

**[0198]** A number of implementations have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the disclosure. For example, various forms of the flows shown above may be used, with steps re-ordered, added, or removed.

**[0199]** All of the functional operations described in this specification may be implemented in digital electronic circuitry, or in computer software, firmware, or hardware, including the structures disclosed in this specification and their structural equivalents, or in combinations of one or more of them. The techniques disclosed may be implemented as one or more computer program products, i.e., one or more modules of computer program instructions encoded on a computer-readable medium for execution by, or to control the operation of, data processing apparatus. The computer readable-medium may be a machine-readable storage device, a machine-readable storage substrate, a memory device, a composition of matter affecting a machine-readable propagated signal, or a combination of one or more of them. The computer-readable medium may be a non-transitory computer-readable medium. The term “data processing apparatus” encompasses all apparatus, devices, and machines for processing data, including by way of example a programmable processor, a computer, or multiple processors or computers. The apparatus may include, in addition to hardware, code that creates an execution environment for the computer program in question, e.g., code that constitutes processor firmware, a protocol stack, a database management system, an operating system, or a combination of one or more of them. A propagated signal is an artificially generated signal, e.g., a machine-generated electrical, optical, or electromagnetic signal that is generated to encode information for transmission to suitable receiver apparatus.

**[0200]** A computer program (also known as a program, software, software application, script, or code) may be written in any form of programming language, including compiled or interpreted languages, and it may be deployed in any form, including as a standalone program or as a module, component, subroutine, or other unit suitable for use in a computing environment. A computer program does not necessarily correspond to a file in a file system. A program may be stored in a portion of a file that holds other programs or data (e.g., one or more scripts stored in a markup language document), in a single file dedicated to the program in question, or in multiple coordinated files (e.g., files that store one or more modules, sub programs, or portions of code). A computer program may be deployed to be executed on one computer or on multiple computers that are located at one site or distributed across multiple sites and interconnected by a communication network.

**[0201]** The processes and logic flows described in this specification may be performed by one or more programmable processors executing one or more computer programs to perform functions by operating on input data and generating output. The processes and logic flows may also be performed by, and apparatus may also be implemented as, special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application specific integrated circuit).

**[0202]** Processors suitable for the execution of a computer program include, by way of example, both general and special purpose microprocessors, and any one or more processors of any kind of digital computer. Generally, a processor will receive instructions and data from a read only memory or a random access memory or both. The essential elements of a computer are a processor for performing instructions and one or more memory devices for storing instructions and data. Generally, a computer will also include, or be operatively coupled to receive data from or transfer data to, or both, one or more mass storage devices for storing data, e.g., magnetic, magneto optical disks, or optical disks. However, a computer need not have such devices. Moreover, a computer may be embedded in another device, e.g., a tablet computer, a mobile telephone, a personal digital assistant (PDA), a mobile audio player, a Global Positioning System (GPS) receiver, to name just a few. Computer readable media suitable for storing computer program instructions and data include all forms of non-volatile memory, media and memory devices, including by way of example semiconductor memory devices, e.g., EPROM, EEPROM, and flash memory devices; magnetic disks, e.g., internal hard disks or removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks. The processor and the memory may be supplemented by, or incorporated in, special purpose logic circuitry.

**[0203]** To provide for interaction with a user, the techniques disclosed may be implemented on a computer having a display device, e.g., a CRT (cathode ray tube) or LCD (liquid crystal display) monitor, for displaying information to the user and a keyboard and a pointing device, e.g., a mouse or a trackball, by which the user may provide input to the computer. Other kinds of devices may be used to provide for interaction with a user as well; for example, feedback provided to the user may be any form of sensory feedback, e.g., visual feedback, auditory feedback, or tactile feedback; and input from the user may be received in any form, including acoustic, speech, or tactile input.

**[0204]** Implementations may include a computing system that includes a back end component, e.g., as a data server, or that includes a middleware component, e.g., an application server, or that includes a front end component, e.g., a client computer having a graphical user interface or a Web browser through which a user may interact with an implementation of the techniques disclosed, or any combination of one or more such back end, middleware, or front end components. The components of the system may be interconnected by any form or medium of digital data communication, e.g., a communication network. Examples of communication networks include a local area network (“LAN”) and a wide area network (“WAN”), e.g., the Internet.

**[0205]** The computing system may include clients and servers. A client and server are generally remote from each other and typically interact through a communication network. The relationship of client and server arises by virtue

of computer programs running on the respective computers and having a client-server relationship to each other.

**[0206]** While this specification contains many specifics, these should not be construed as limitations, but rather as descriptions of features specific to particular implementations. Certain features that are described in this specification in the context of separate implementations may also be implemented in combination in a single implementation. Conversely, various features that are described in the context of a single implementation may also be implemented in multiple implementations separately or in any suitable subcombination. Moreover, although features may be described above as acting in certain combinations and even initially claimed as such, one or more features from a claimed combination may in some cases be excised from the combination, and the claimed combination may be directed to a subcombination or variation of a subcombination.

**[0207]** Similarly, while operations are depicted in the drawings in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system components in the implementations described above should not be understood as requiring such separation in all implementations, and it should be understood that the described program components and systems may generally be integrated together in a single software product or packaged into multiple software products.

**[0208]** Thus, particular implementations have been described. Other implementations are within the scope of the following claims. For example, the actions recited in the claims may be performed in a different order and still achieve desirable results.

What is claimed is:

1. A method implemented by one or more processors, the method comprising:

receiving user input from a human user to initiate performance of a task using a bot, wherein performance of the task using the bot comprises:

simultaneously initiating a plurality of telephone calls to corresponding organizations, and

simultaneously conducting a corresponding telephone conversation between a corresponding instance of the bot and a corresponding human representative of each of the corresponding organizations;

identifying, based on the user input, at least one value for a corresponding parameter utilized while conducting the corresponding telephone conversations;

generating, based on corresponding additional user input received from the corresponding human representative of each of the corresponding organizations, corresponding output associated with information related to the at least one value for the corresponding parameter for each of the corresponding organizations; and

causing, while simultaneously conducting a corresponding telephone conversations, the corresponding output associated with the information related to the at least one value for the corresponding parameter for each of the corresponding organizations to be provided for presentation to the human user via a computing device of the human user.

2. The method of claim 1, further comprising:

in response to causing the corresponding output to be provided for presentation to the human user:

receiving a human user selection from the human user of given output associated with a given organization of the corresponding organizations, and  
causing the bot to complete performance of the task with the given organization.

3. The method of claim 2, wherein causing the bot to complete performance of the task with the given organization comprises:

causing the bot to utilize the at least one value for the corresponding parameter in furtherance of the corresponding telephone conversation between the corresponding instance of the bot and the corresponding human representative of the given organization.

4. The method of claim 3, wherein the task comprises one of: an appointment task, a reservation task, an inventory task, or a transportation task.

5. The method of claim 2, further comprising:

causing the bot to terminate the corresponding telephone conversations with the corresponding human representatives of the other corresponding organizations.

6. The method of claim 2, wherein the corresponding output associated with the information related to the at least one value for the corresponding parameter for each of the corresponding organizations provided for presentation to the human user is provided via a graphical user interface of the computing device of the human user, and wherein receiving the human user selection from the human user of the given output comprises receiving the human user selection via the graphical user interface of the computing device.

7. The method of claim 2, wherein the corresponding output associated with the information related to the at least one value for the corresponding parameter for each of the corresponding organizations provided for presentation to the human user is provided via a one or more speakers of the computing device of the human user, and wherein receiving the human user selection from the human user of the given output comprises receiving the human user selection via microphones of the computing device.

8. The method of claim 1, wherein the corresponding organizations are identified based on further additional user input received from the human user prior to initiating the plurality of telephone calls to the corresponding organizations.

9. The method of claim 1, wherein the corresponding organizations are identified based on a current location of the computing device of the human user.

10. The method of claim 1, wherein the user input received from the human user to initiate performance of the task is received via a search interface of the computing device of the human user.

11. A system comprising:

one or more processors; and

memory storing instructions that, when executed, cause one or more of the processors to:

receive user input from a human user to initiate performance of a task using a bot, wherein performance of the task using the bot comprises:

simultaneously initiate a plurality of telephone calls to corresponding organizations, and

simultaneously conduct a corresponding telephone conversation between a corresponding instance of

the bot and a corresponding human representative of each of the corresponding organizations;  
 identify, based on the user input, at least one value for a corresponding parameter utilized while conducting the corresponding telephone conversations;  
 generate, based on corresponding additional user input received from the corresponding human representative of each of the corresponding organizations, corresponding output associated with information related to the at least one value for the corresponding parameter for each of the corresponding organizations; and  
 cause, while simultaneously conducting a corresponding telephone conversations, the corresponding output associated with the information related to the at least one value for the corresponding parameter for each of the corresponding organizations to be provided for presentation to the human user via a computing device of the human user.

**12.** The system of claim **11**, wherein the instructions further cause one or more of the processors to:  
 in response to causing the corresponding output to be provided for presentation to the human user:  
 receive a human user selection from the human user of given output associated with a given organization of the corresponding organizations, and  
 cause the bot to complete performance of the task with the given organization.

**13.** The system of claim **12**, wherein the instructions to cause the bot to complete performance of the task with the given organization comprise instructions to:  
 cause the bot to utilize the at least one value for the corresponding parameter in furtherance of the corresponding telephone conversation between the corresponding instance of the bot and the corresponding human representative of the given organization.

**14.** The system of claim **13**, wherein the task comprises one of: an appointment task, a reservation task, an inventory task, or a transportation task.

**15.** The system of claim **12**, wherein the instructions further cause one or more of the processors to:  
 cause the bot to terminate the corresponding telephone conversations with the corresponding human representatives of the other corresponding organizations.

**16.** The system of claim **12**, wherein the corresponding output associated with the information related to the at least one value for the corresponding parameter for each of the corresponding organizations provided for presentation to the human user is provided via a graphical user interface of the computing device of the human user, and wherein the instructions to receive the human user selection from the

human user of the given output comprise instructions to receive the human user selection via the graphical user interface of the computing device.

**17.** The system of claim **12**, wherein the corresponding output associated with the information related to the at least one value for the corresponding parameter for each of the corresponding organizations provided for presentation to the human user is provided via a one or more speakers of the computing device of the human user, and wherein the instructions to receive the human user selection from the human user of the given output comprise instructions to receiving the human user selection via microphones of the computing device.

**18.** The system of claim **11**, wherein the corresponding organizations are identified based on further additional user input received from the human user prior to initiating the plurality of telephone calls to the corresponding organizations.

**19.** The system of claim **11**, wherein the corresponding organizations are identified based on a current location of the computing device of the human user.

**20.** A non-transitory computer-readable storage medium storing instructions that, when executed, cause one or more processors to:

receive user input from a human user to initiate performance of a task using a bot, wherein performance of the task using the bot comprises:

simultaneously initiate a plurality of telephone calls to corresponding organizations, and

simultaneously conduct a corresponding telephone conversation between a corresponding instance of the bot and a corresponding human representative of each of the corresponding organizations;

identify, based on the user input, at least one value for a corresponding parameter utilized while conducting the corresponding telephone conversations;

generate, based on corresponding additional user input received from the corresponding human representative of each of the corresponding organizations, corresponding output associated with information related to the at least one value for the corresponding parameter for each of the corresponding organizations; and

cause, while simultaneously conducting a corresponding telephone conversations, the corresponding output associated with the information related to the at least one value for the corresponding parameter for each of the corresponding organizations to be provided for presentation to the human user via a computing device of the human user.

\* \* \* \* \*